Theory of Principal Component Filter Banks with Applications to Multicomponent Imagery

Theory of Principal Component Filter Banks with Applications to Multicomponent Imagery

by

Mihaela Dobre Pal, B.S., M.S.

DISSERTATION

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT AUSTIN

August 2003

In memory of my mother, Maria.

Acknowledgments

Theory of Principal Component Filter Banks with Applications to Multicomponent Imagery

Mihaela Dobre Pal, Ph.D.
The University of Texas at Austin, 2003

Supervisors: E. Ward Cheney
                Christopher M. Brislawn

In the first part of the thesis we give background about the digital signal processing, required throughout. We introduce the Karhunen-Loève transform and the most commonly used optimality criteria for orthonormal uniform filter banks.

In the second part of the thesis the definition of principal component filter banks is given; these filter banks unify the theory of optimality of filter banks under explicitly stated criteria. We discuss the existence of principal component filter banks and present a study case pertaining to autoregressive input signals and finite impulse response filter banks. We prove a theorem on the existence of coding gain optimal finite impulse response filter banks. For filter banks with two channels, coding gain optimal filter banks are also principal component filter banks.

As an application of the theory of optimal filter banks we design two-channel principal component filter banks for remote sensing hyper-spectral

images. These filter banks are used to decorrelate an image, i.e. to represent the image in a more compact form. This design strategy leads do a more efficient compression of large images within the JPEG-2000 paradigm.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1   Motivation

This work on principal component filter banks for multicomponent imagery was initiated in the context of developing extensions for JPEG-2000 [2, 3], the latest image compression paradigm added to a wide collection of international standards developed by the Joint Photographic Experts Group (JPEG). Compression of images with JPEG-2000 is a response to higher demands from modern multi-media communication, internet traffic and many similar applications where compression is indispensable. The compression performance of the new standard is superior to the previous JPEG [109, 89] and a new class of extensions is targeted at the efficient compression and manipulation of large collection of images named "multicomponent" images [22].

Multicomponent images are two-dimensional images stacked together, just as the familiar three mono-color red, blue and green images that render the usual color photograph; two coordinates are the usual plane coordinates and the third direction, the so-called "component" direction, characterizes the type of image. One class of multicomponent images comprises multi-spectral and hyper-spectral images – like remote sensing data. These images can be

regarded as a refinement of a color photograph with more wavelengths at which images are recorded. Multi-spectral images have a few dozen images and hyper-spectral images have hundreds – even thousands – of images [99]. The component direction is in this case the wavelength, or the spectral direction. We can imagine a hyper-spectral image as a cube such as the one in Figure 1.1 showing Los Alamos.
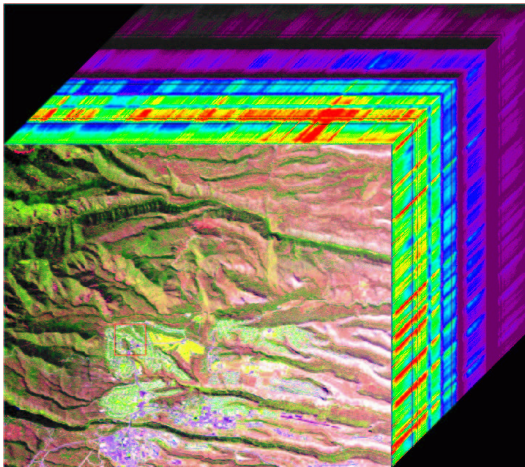
Figure 1.1: Los Alamos hyper-spectral cube

Other major types of multicomponent imagery are 3-D volumetric data, where the component direction is the third Euclidean coordinate. These images are tomograms – also known as computer assisted tomography (CAT) scans – and magnetic resonance images from scientific experiments or from medical investigations [1, 126, 69].

Multicomponent imagery requires very large storage and long transmission duration; therefore, compression of such data sets is very desirable.

There are different compression techniques depending on the type of data [98, 110, 116, 20].

Compression is of two kinds: lossy and lossless. With lossy compression there is loss of accuracy of the reconstructed data, or distortion, whereas with lossless compression, the reconstructed images coincide with the original uncompressed ones. Both types of compression are widely used, depending on the data processing tasks. JPEG-2000 gives a better resolution at low bit rates than its previous counterpart [23, 72, 111].

Ultimately, our work is aimed at efficient compression of multicomponent images, particularly, hyper-spectral remote sensing images. The compression mechanism is data dependent and images without redundancy – or, technically speaking, *uncorrelated* images – are subject to a more efficient compression [40].

Hyper-spectral images are large data sets, on the order of hundreds of megabytes [123]. There is a high correlation in the spectral direction for most remote sensing multi-spectral and hyper-spectral images [27, 35].

Our task is to investigate the existence and design of tools that decorrelate these images so that the data is represented in a more compact form. When a hyper-spectral image is compressed in a lossy manner, this advantage of data reduction should be used in applications where there is little distortion compared to the same tasks performed on uncompressed images [59, 102].

We expect that compression of highly correlated hyper-spectral images

is more efficient when decorrelation in the spectral direction is applied prior to compression, and the experiments on real data show that this is true. First we employ the Karhunen-Loève transform – also called principal component transform – which is known to be one of the the best decorrelators in image processing. This transform belongs to a small class of image and general signal processing tools, namely, the class of orthogonal transforms. We also use a wavelet transforms for decorrelation in the spectral direction [86]. It is desirable to find effective decorrelators among more general signal processing tools, namely filters with memory that exploits the statistical properties of the data.

A decorrelation tool generates a more compact representation of a signal and packs the energy of a signal into a few information carriers. An example is to use only a few, out of the hundreds of two-dimensional images from a hyper-spectral data set, and have a minimum loss of features or information for a certain image analysis task. The mathematical answer is, in general, given in terms of principal components. Besides the Karhunen-Loève transform, there are more sophisticated signal processing tools that achieve similar decorrelation; in short, these are the principal component filter banks.

Theoretical results on principal component filter banks leave the existence issue still open for most classes of realizable filters and for arbitrary input signals. These filter banks have the attractive quality of unifying the usual optimality criteria for filter banks. The quest for this type of filter remains an ongoing task. Our starting point is to study feature extraction from com-

4

pressed images, and this challenging task constantly motivates us to search for optimal filter banks and for principal component filter banks – the "ultimate" optimal filter banks.

Our present and future research comprises two directions. One is the existence and design of filter banks that are optimal under explicitly stated criteria. The other aim of our work is to design two-channel principal component filter banks that can be embedded in the JPEG-2000 Standard. The rational for this is to reduce the cost of the high-dimension Karhunen-Loève transform. A natural continuation of this research also targets a comprehensive statistical approach to the existence of principal component filter banks. The existence of principal component filter banks for general classes of input signals and realizable classes of filters is still open; however, it is of great interest to know if there might be particular classes of inputs and categories that have principal component filter banks.

1.2   Preliminaries on Digital Signal Processing

Today's demands on information exchange, communications, and analysis couple typical engineering problems with the mathematical research from many areas: statistics, functional analysis, information theory, systems theory, approximation theory, numerical analysis, and harmonic analysis – to name a few. Among the variety of experts working in digital signal processing, electrical engineers and mathematicians form the majority. This section presents, in a unified manner, the notation and the elementary processing blocks that are

the foundations of more complex mathematical and engineering studies and practical solutions.

1.2.1 Notation

The notation used here follows Mallat [72], Strang and Nguyen [106] and Vaidyanathan [117].

- The conjugate of a complex number z is denoted by z^*.

- Let $\boldsymbol{E}(\cdot)$ be a matrix-valued function of a complex variable z.

  - The transpose is denoted by $\boldsymbol{E}^T$ and $\boldsymbol{E}^T(z) = [\boldsymbol{E}(z)]^T$.

  - The conjugate transpose of $\boldsymbol{E}$ is denoted by a dagger: $\boldsymbol{E}^\dagger(z) = [\boldsymbol{E}(z)]^{*T}$.

  - The conjugate of the coefficients only, not the argument, is denoted by: $\boldsymbol{E}_*(z) = [\boldsymbol{E}(z^*)]^*$

  - The tilde notation is used for *para-conjugation*: $\widetilde{\boldsymbol{E}}(z) = \boldsymbol{E}_*^T(z^{-1})$ for all complex z. Observe that $\widetilde{\boldsymbol{E}}(z) = \boldsymbol{E}^\dagger(1/z^*)$ for any z. On the unit circle we have: $\widetilde{\boldsymbol{E}}(z) = \boldsymbol{E}^\dagger(z)$.

- The n-th term of a sequence $w : \mathbb{Z} \to \mathbb{C}$ is denoted by $w[n], n \in \mathbb{Z}$.

- The n-th term of a vector sequence $\boldsymbol{x}$ is denoted by $\boldsymbol{x}[n]$ and its k-th component by $\boldsymbol{x}_k[n], \ k, n \in \mathbb{Z}, k \geq 0$.

- In all equations, formulae and relations the symbol i denotes exclusively the square root of -1.

6

1.2.2 Signals

A *signal* is a function that communicates information about the state or behavior of a physical system [84, 85]. Digital signal processing is widely used in a variety of fields such as: acoustics, sonar radar, biomedical engineering, seismology, speech, imagery, and other types of data communication [14, 70, 75]. Signals are represented mathematically as functions of one or more independent variables. The independent variable(s) may be continuous or discrete and the functions that represent continuous signals are of continuous variables, whereas the discrete signals are represented by functions defined on discrete sets [30, 72].

In most applications a discrete signal is regarded as the discretization of a continuous time signal realized by sampling an analog signal at certain discrete values. Discrete signals are the subject of digital signal processing.

Discrete signals are denoted by x, y or $\{x[n]\}_{n \in \mathbb{Z}}, \{y[n]\}_{n \in \mathbb{Z}}$ with the meaning that x and y are functions defined on the set of integer numbers, $\mathbb{Z}$, and at each "temporal" value $x[n]$, n denotes the *time index*. Sequences describing signals are complex-valued unless otherwise specified. We use the terms "signal" and "sequence" interchangeably.

Two signals often encountered are

- The unit-pulse.

$$\delta[n] = \begin{cases} 1, & n = 0 \\ 0 & \text{otherwise} \end{cases} \tag{1.1}$$

7

- The unit-step:

$$U[n] = \begin{cases} 1, & n \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

1.2.2.1 Transformed Versions of Signals

The frequency domain representation of a signal is used in many studies. The most common transforms are the z-transform and the discrete Fourier transform. Let x be a signal; then its z-transform $X(\cdot)$, is defined as

$$X(z) = \sum_{n \in \mathbb{Z}} x[n] z^{-n}, \qquad z \in \mathbb{C}$$

The discrete Fourier transform $X(\cdot)$ is the z-transform evaluated on the unit circle $K(0,1) = \{z = e^{i\omega}, w \in [0, 2\pi]\}$. To show this connection between the z-transform and the Fourier transform, $e^{i\omega}$ is used as the argument at which the Fourier transform is evaluated. Properties of the z transform and the Fourier transforms can be found in [41, 25, 84, 85].

1.2.2.2 Signal Models

As in many other areas of mathematics, the models for signals are chosen with the purpose of developing theory well-suited for practical applications [33, 101, 77]. For example, a model that is expressed in simple mathematical terms can be used to design and implement a decorrelating tool that reflects the properties of the input. In practice, when we observe signals with similar characteristics as the model, then we have a tool that is immediately available.

Let $x : \mathbb{Z} \to \mathbb{C}$ be a discrete time stochastic process. At each value of the integer argument n, $x[n]$ is a random variable. Equivalently, we can say

that a signal is a sequence of random variables. Denote by $E\{x[n]\}$ the mean of the random variable $x[n]$. Define the autocorrelation R_{xx} sequence as

$$R_{xx}[n, m] = E\{x[n]x^*[m]\}, \qquad \text{for all } m, n \in \mathbb{Z}$$

A process is *wide-sense stationary* if 1) the mean is constant $E\{x[n]\}$ for all integers n and 2) the autocorrelation $R_{xx}[m, n] = E\{x[n]x^*[m]\}$, depends only on the *lag*, $m - n$ [88], which is to say that

$$R_{xx}[n, m] = R_{xx}[n - m] = E\{x[n]x^*[m]\}, \qquad \text{for all } m, n \in \mathbb{Z}. \qquad (1.2)$$

We consider only processes that are wide-sense stationary and also have a zero mean. The variance – or energy – of a zero mean process is

$$\sigma_x^2 = R_{xx}[0] = E\{x[n]x^*[n]\}. \qquad (1.3)$$

We consider only processes with finite variance. In practice all signals have finite length; however most theoretical results assume processes defined on the whole set of integers. From the autocorrelation sequence we determine the power spectral density of the input via a discrete Fourier transform,

$$S_{xx}(e^{i\omega}) = \sum_{k \in \mathbb{Z}} R_{xx}[k]e^{-i\omega k},$$

Here we assume that the correlation sequence is square summable. It can be shown that the power spectral density is positive [33, 88, 38, 37]. A real process has a symmetric autocorrelation sequence.

Let M be a nonnegative integer. A *blocked signal* is a vector sequence $\boldsymbol{x} : \mathbb{Z} \to \mathbb{C}^M$ with components

$$\boldsymbol{x}[n] = [\ x[nM] \quad x[nM - 1] \quad \dots \quad x[nM - M + 1]\ ]^T \qquad (1.4)$$

9

The blocked signal is formed with every sequential subset of M elements of the original signal. A blocked signal is a discrete time vector random process whose components are themselves discrete random processes. Imposing the wide-sense stationarity condition on the components of the blocked signal is equivalent to the assumption that the signal is wide-sense cyclo-stationary (see [97]). A wide sense stationary process x is said to be cyclo-stationary with period M if for all integers k and n

$$E\{x[n]\} = E\{x[n + kM]\},$$
$$R_{xx}[n, k] = R_{xx}[n + M, k].$$

For a complex stochastic process, stationarity is defined in the sense of joint stationarity of the real and imaginary parts of the process [30].

Let $\{\boldsymbol{R_{xx}}[k]\}_{k \in \mathbb{Z}}$ be the autocorrelation matrix sequence of the blocked version, $\boldsymbol{x}$, given by

$$\boldsymbol{R_{xx}}[k] = E\{\boldsymbol{x}[n]\,\boldsymbol{x}^{\dagger}[n - k]\}, \qquad \text{for all } n \in \mathbb{Z}$$

At lag $k = 0$ the matrix $\boldsymbol{R_{xx}}[0]$ is a real symmetric, Toeplitz matrix:

$$\boldsymbol{R_{xx}}[0] = \begin{bmatrix} R[0] & R[1] & R[2] & \dots & R[M-2] & R[M-1] \\ R[1] & R[0] & R[1] & \dots & R[M-3] & R[M-2] \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ R[M-1] & R[M-2] & R[M-3] & \dots & R[1] & R[0] \end{bmatrix}$$

The power spectral density matrix is the discrete Fourier transform of the autocorrelation matrix sequence

$$\boldsymbol{S_{xx}}(e^{i\omega}) = \sum_{k \in \mathbb{Z}} \boldsymbol{R_{xx}}[k] e^{-i\,\omega\,k} \tag{1.5}$$

10

We list the constraints imposed on the models for the input signals throughout. We consider discrete stochastic processes that

- have finite energy,

- have a mean of zero,

- have correlation sequence square summable,

- have power spectral density vanishing on at most a set of measure zero,

- whose blocked versions are wide-sense stationary.

We now introduce a model of autoregressive signals. Many real experiments are modeled by autoregressive sequences and we use an autoregressive signal in Chapter 4.

A discrete-time stochastic process $x : \mathbb{Z} \to \mathbb{C}$ is autoregressive if it is described by

$$x[n] = \rho x[n-1] + w[n], \qquad n \in \mathbb{Z}$$

where $w : \mathbb{Z} \to \mathbb{C}$ is a Gaussian white noise process with variance σ_w^2 and the *correlation coefficient* satisfies $|\rho| < 1$. This is a first-order Markov process; it is also called a Gauss-Markov – or autoregressive(1) – process and it is denoted by AR(1); see [40, 75]. The noise mean satisfies $E\{w[n]\} = 0$, for any integer n, and the autocorrelation of the noise is

$$R_{ww}[n, k] = E\{w[n]w[k]^*\} = \sigma_w^2 \delta_{nk}, \qquad n, k \in \mathbb{Z}.$$

The variances of the autoregressive AR(1) process, σ_x^2, and that of the noise are related by

$$\sigma_x^2 = \frac{\sigma_w^2}{1 - \rho^2}.$$

The mean of the input process x is zero

$$\mu_x = E\{x[n]\} = 0, \qquad n \in \mathbb{Z}.$$

The autocorrelation of the autoregressive AR(1) process at lag k is $R_{xx}[k] = \rho^k \sigma_x^2$, and it is symmetric

$$R_{xx}[k] = R_{xx}[-k] = \rho^{|k|} \sigma_x^2.$$

The power spectral density of the process is

$$S_{xx}(e^{i\omega}) = |T(e^{i\omega})|^2 S_{ww}(e^{i\omega}), \text{ where } T(e^{i\omega}) = 1/(1 - \rho e^{-i\omega}).$$

Consider a blocked version of the signal, described by Equation (1.4). Without loss of generality the input variance can be normalized so that $\sigma_x^2 = 1$. The autocorrelation at lag zero is

$$\boldsymbol{R_{xx}}[0] = \begin{bmatrix} 1 & \rho & \cdots & \rho^{M-1} \\ \rho & 1 & \rho & \cdots \\ \vdots & & & \\ \rho^{M-1} & \cdots & \rho & 1 \end{bmatrix}$$

Denote

$$\boldsymbol{R_M}(\rho) = \begin{bmatrix} 1 & \rho & \cdots & \rho^{M-1} \\ \rho^{-1} & 1 & \rho & \cdots \\ \vdots & & & \\ \rho^{-M+1} & \cdots & \rho^{-1} & 1 \end{bmatrix}$$

The autocorrelation at non-zero lags is

$$\boldsymbol{R_{xx}}[k] = \begin{cases} \rho^{Mk}\boldsymbol{R}_M(\rho), & k > 0 \\ \rho^{-Mk}\boldsymbol{R}_M(\rho)^T, & k < 0 \end{cases}$$

and it satisfies $\boldsymbol{R_{xx}}[-k] = \boldsymbol{R}_{\boldsymbol{xx}}^{\dagger}[k]$.

The power spectral density matrix, by Equation (1.5), is:

$$\begin{aligned} \boldsymbol{S_{xx}}(z) &= \boldsymbol{R_{xx}}(0) + \sum_{k=1}^{\infty}\boldsymbol{R_{xx}}(k)z^{-k} + \sum_{k=1}^{\infty}\boldsymbol{R_{xx}}(-k)z^{k} && (1.6) \\ &= \boldsymbol{R_{xx}}(0) + \boldsymbol{R}_M(\rho)\sum_{k=1}^{\infty}\rho^{Mk}z^{-k} + \boldsymbol{R}_M(\rho)^T\sum_{k=1}^{\infty}\rho^{Mk}z^{k} \\ &= \boldsymbol{R_{xx}}(0) + \frac{\rho^{M}z^{-1}}{1 - \rho^{M}z^{-1}}\boldsymbol{R}_M(\rho) + \frac{\rho^{M}z}{1 - \rho^{M}z}\boldsymbol{R}_M(\rho)^T \end{aligned}$$

where $|\rho| < 1$ and $|z| = 1$.

Autoregressive processes are used to model real applications and are also used in theoretical studies in signal processing [11, 75, 16].

1.3   Systems and Transfer Functions

Given a stochastic process x we can create another process y by a mapping H:

$$y = H(x),$$

We say that y is the output of the *system H* with input x. We call H a *single-input single-output* system if its input and output are one-dimensional (scalar) processes. If both the input and output are vector random processes the system is called *multi-input multi-output* [88, 117, 101].
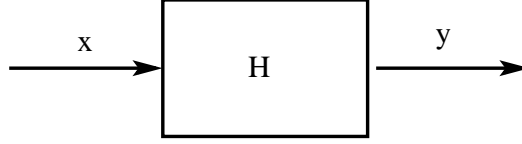
Figure 1.2: A system H acting on an input signal x

A *discrete time system* operates on input $\{x[n]\}_{n\in\mathbb{Z}}$ to produce $\{y[n]\}_{n\in\mathbb{Z}}$

$$y[n] = H(x[n]), \qquad n \in \mathbb{Z}$$

A few properties and characteristics of discrete-time systems follow [117]:

- A system is *memoryless* if its output at time n depends on the input at time n only and not on past or future values of the input.

- Suppose the input sequences $\{x[n]\}_{n\in\mathbb{Z}}$ and $\{u[n]\}_{n\in\mathbb{Z}}$ produce the output sequences $\{y[n]\}_{n\in\mathbb{Z}}$ and $\{v[n]\}_{n\in\mathbb{Z}}$ through the system H. If the output from $\alpha x + \beta u$ is $\alpha y + \beta v$, for all α, β, x, u, then we say that the system is *linear*.

- When the output from $x[n-k]$ is $y[n-k]$ for all integers k, n, and for any signal x, we say that H is a *shift invariant* or *translation invariant system* or *time invariant*.

- A system is called *linear time (or translation) invariant* if it is both linear and time (or shift) invariant. We also call such a system a *filter*.

- A linear translation invariant system H is characterized by its *impulse response* sequence, $\{h[n]\}_{n\in\mathbb{Z}}$, which is the output of the system H on

14

the unit-pulse δ defined in Equation (1.1). The impulse response is also called the *unit-pulse* response.

- A discrete time system is called *causal* if the output $y[n]$ at time n does not depend on future values of the input sequence $x[m]$, with $m > n$, $m, n \in \mathbb{Z}$. This condition is satisfied by a linear translation invariant system if and only if the impulse response satisfies $h[n] = 0$ for all $n < 0$. We also call a sequence x causal if its terms satisfy $x[n] = 0$ for all $n < 0$. An anti-causal sequence is one whose terms satisfy: $x[n] = 0$ for all $n \geq 0$.

- A system is called *real* if the output is real whenever the input is real. For linear time invariant systems this is equivalent to the condition that the impulse response $\{h[n]\}_{n \in \mathbb{Z}}$ is a real valued sequence.

- A discrete system is *stable* if any bounded input produces a bounded output. For linear time invariant systems, the bounded input bounded output condition is equivalent to the absolute summability of the impulse response

$$\sum_{n \in \mathbb{Z}} |h[n]| < \infty \, .$$

For linear time invariant systems, the input-output relation is the convolution

$$y[n] = \sum_{k \in \mathbb{Z}} h[k]x[n-k] \, . \tag{1.7}$$

15

In the z-transform domain we have

$$Y(z) = H(z)X(z), \qquad\qquad (1.8)$$

where $H(\cdot)$ is the z-transform of h, i.e. $H(z) = \sum_{n\in\mathbb{Z}} h[n]z^{-n}$ and $H(\cdot)$ is called the *transfer function* of the linear time invariant system. A transfer function can be equivalently represented as a recursive difference equation of the form

$$b_0 y_n = -\sum_{m=1}^{N} b_m y_{n-m} + \sum_{m=0}^{N} a_m x_{n-m}.$$

A system is causal if $b_0 \neq 0$ and we can assume without loss of generality that $b_1 = 1$.

We only consider systems with *rational transfer functions* of the form

$$H(z) = \frac{A(z)}{B(z)} \qquad \text{with} \qquad\qquad (1.9)$$

$$A(z) = \sum_{n=0}^{N} a_n z^{-n}, \ \ B(z) = \sum_{n=0}^{N} b_n z^{-n}, \qquad a_n, b_n \in \mathbb{C}.$$

When $A(z)$ and $B(z)$ are relatively prime we say that the system H is in irreducible rational form and N is the *order* of the system.

1.3.0.3 *Finite impulse response and infinite impulse response systems*

A finite impulse response system is such that in its transfer function there is only one index n for which b_n is nonzero (see Equation 1.9): there exists a unique n_0 such that $b_{n_0} \neq 0$ and $b_n = 0$ for all $n \neq n_0$. Example: $H(z) = a_0 z^2 + a_1 z + a_2 + a_3 z^{-1}$ is a finite impulse response system.

If a finite impulse response causal system $H(z) = \sum_{n=0}^{N} h[n]z^{-n}$ has the coefficient $h[N]$ nonzero we say that N is the *order* of the system and that the length of the filter is $N + 1$. If a system is not a finite impulse response system then it is an infinite impulse response system. Obviously, the length of an infinite impulse response filter is infinite. An example of an infinite impulse response filter is:

$$H(z) = \frac{1}{1 - az^{-1}} .$$

A finite impulse response system is also called *all-zero* [117, 87].

1.3.1   Digital filters

Earlier we introduced a filter as a linear time-invariant system [106, 117]. Figure 1.2 shows a filter as well when H is a discrete linear translation invariant system. The input x is passed through the filter $H(\cdot)$ that renders the output y; in the time domain and in the frequency domain the input-output relations are, as in Equations (1.7) and (1.8):

$$
\begin{aligned}
y[n] &= \sum_{k \in \mathbb{Z}} h[k]\, x[n - k] \\
Y(z) &= H(z)\, X(z)
\end{aligned}
$$

1.3.2   Polyphase representation of a filter

Consider a filter with impulse response $\{h[n]\}_{n \in \mathbb{Z}}$ and frequency response:

$$H(z) = \sum_{n \in \mathbb{Z}} h[n]z^{-n} .$$

We show first the two-channel polyphase representation. The previous equation can be written equivalently as

$$H(z) = \sum_{n \in \mathbb{Z}} h[2n] z^{-2n} + z^{-1} \sum_{n \in \mathbb{Z}} h[2n+1] z^{-2n} \,.$$

Define

$$E_0(z) = \sum_{n \in \mathbb{Z}} h[2n] z^{-n}$$

and

$$E_1(z) = \sum_{n \in \mathbb{Z}} h[2n+1] z^{-n}$$

so that $H(z)$ can be represented as

$$H(z) = E_0(z^2) + z^{-1} E_1(z^2) \,.$$

This is a representation with two phases. In general, for a M-channel filter bank we have [117, 107]

$$H(z) = \begin{cases} \sum_{n \in \mathbb{Z}} h[nM] z^{-nM} \\ + z^{-1} \sum_{n \in \mathbb{Z}} h[nM+1] z^{-nM} \\ \vdots \\ + z^{-(M-1)} \sum_{n \in \mathbb{Z}} h[nM+M-1] z^{-nM} \end{cases}$$

or

$$\begin{aligned} H(z) &= \sum_{l=0}^{M-1} z^{-l} E_l(z^M) \,, \\ E_l(z) &= \sum_{n \in \mathbb{Z}} e_l[n] z^{-n} \,, \\ e_l[n] &= h[Mn+l], 0 \le l < M \,. \end{aligned}$$

(1.10)

18

1.3.3 Basic signal processing blocks

A digital signal processing system that operates on a discrete blocked signal is called a multi-rate system. The three main building blocks that are part of any multi-rate system are: the linear time invariant filters, the down-sampler and the up-sampler. An M-fold down-sampler keeps only every M-th element of a signal sequence and can be regarded as the constructor of the components of the blocked version of a signal. Mathematically, this is expressed, in the time domain and in the frequency domain, as

$$
\begin{aligned}
y[n] &= x[nM], \qquad n \in \mathbb{Z}, \\
Y(z) &= (X(z))|_{\downarrow M} = \frac{1}{M} \sum_{k=0}^{M-1} X(z^{1/M} e^{-i\frac{2k\pi}{M}}).
\end{aligned}
$$

The M-fold up-sampler inserts $M-1$ zeros between consecutive samples of a signal; the output is

$$
\begin{aligned}
y[n] &= \begin{cases} x[n/M], & \text{if } n/M \in \mathbb{Z} \\ 0 & \text{otherwise} \end{cases} \\
Y(z) &= (X(z)) \uparrow M = X(z^M).
\end{aligned}
$$

The effect of down-sampling is aliasing: in the frequency domain the output from a down-sampler consists of an M-fold expansion of the input in the frequency domain, as in Figure 1.3 (a). We call X an *alias-free(M)* or *anti-alias(M)* signal if these shifted copies do not overlap (or alias). An alias-free signal is said to have alias-free(M) support. We define the M *alias frequencies* of a frequency ω by:

$$
\omega_k = \omega + \frac{2k\pi}{M}, \qquad \omega \in [0, 2\pi], \quad k = 0, 1, \ldots, M-1
$$

19

a) Downsampling and aliasing   b) Upsampling and imaging
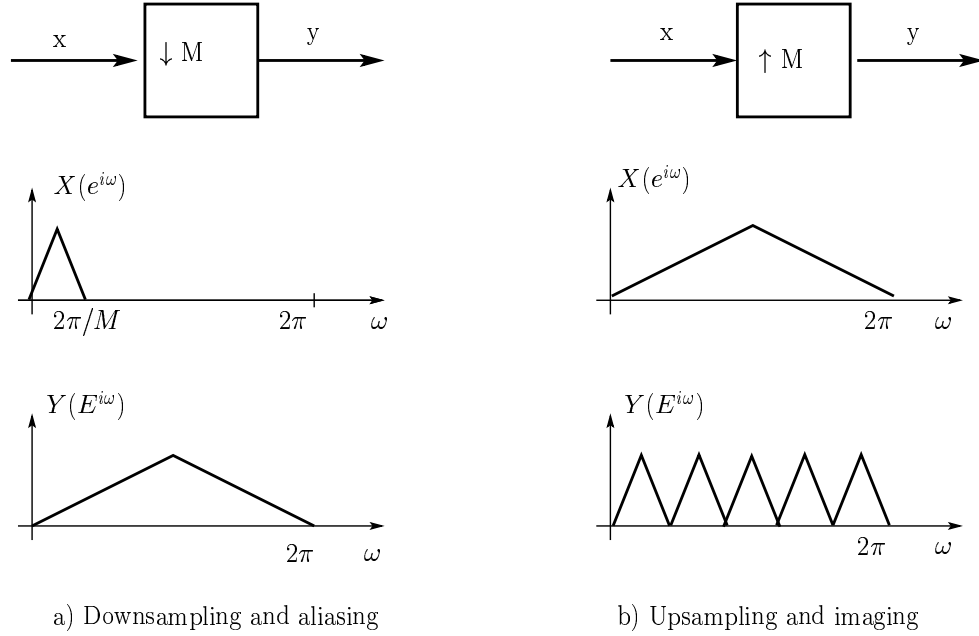
Figure 1.3: Down-sampler and up-sampler

and express the alias-free (M) condition in terms of the alias frequencies as: X is alias-free(M) if $X(e^{i\omega_k})$ is nonzero for at most one index k.

The effect of the up-sampler is called imaging and it means that M copies of the original with an M-fold compression in the frequency domain are produced. Figure 1.3(b) shows the up-sampler and its effects in the frequency domain.

A particular infinite impulse response type of filter will often be encountered in the subsequent chapters. It is the *ideal passband filter* whose frequency response is a step function as shown in Figure 1.4. This type of filter is unrealizable and is sometimes called an "ideal" filter. The low-pass ideal filter has the nonzero response – or the passband – in $[0, \pi/2) \cup [3\pi/2, 2\pi]$

20
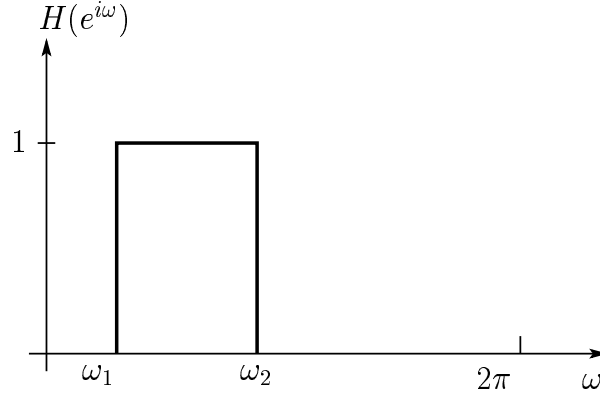
Figure 1.4: Ideal passband filter

and a high-pass correspondent has the frequency response in the remainder of the spectrum. There are also ideal filters with multiple pass-bands where the impulse response is 1 on a collection of intervals in $[0, 2\pi]$.

1.3.4 Filter banks

A scheme such as the one in Figure 1.5 is called a filter bank. It is a collection of filters, down-samplers and up-samplers acting on an input signal x. The filters $H_0, H_1, \ldots, H_{M-1}$ form the *analysis* filter bank; the *synthesis* filter bank consists of the filters $F_0, F_1, \ldots, F_{M-1}$. The input signal x is sent through the analysis filters that output processes $x_k, k = 0, 1, \ldots, M-1$. The M-fold down-samplers render the signals v_k's. The dots in the figure represent possible intermediate processing tasks, such as transmission, compression; some of these tasks may induce distortion, in which case we simply call these processes simply quantizers (see Section 1.4). After processing, the signal is up-sampled, resulting in the processes u_k that are filtered by the filters

F_k, $k = 0, 1, \ldots, M - 1$. Finally the output signal $\widehat{x}$ is formed by summing the synthesized components.
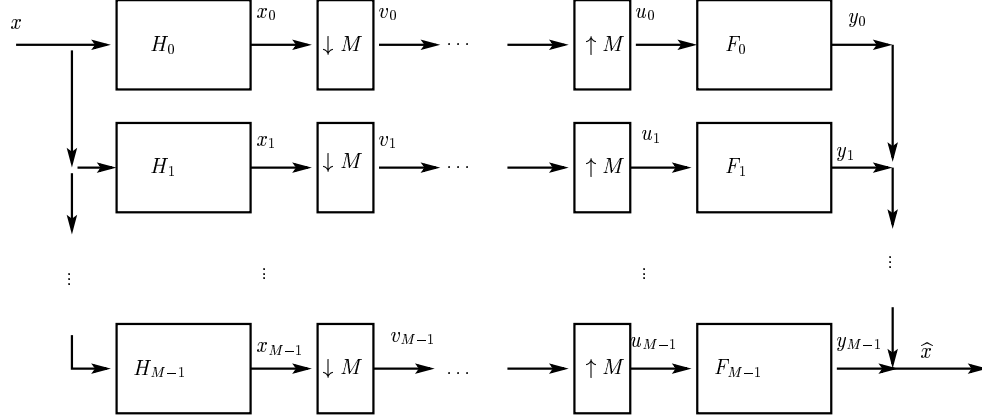
Figure 1.5: Standard form of a uniform maximally down-sampled M-channel filter bank.

A filter bank has the perfect reconstruction property when the output is a scaled and shifted version of the input, i.e., there exists an integer l such that, for all integers n:

$$\hat{x}[n] = Cx[n - l], \qquad C \in \mathbb{C} \setminus \{0\}. \tag{1.11}$$

Values of 1 for C and of zero for l are possible in a perfect reconstruction filter bank. Perfect reconstruction assumes no loss of information (distortion). Therefore, in a setting with a perfect reconstruction filter bank and intermediate processing tasks the error comes only from the intermediate processing and the distortion is viewed as quantization noise.

1.4 Coding Theory and Quantization

Signals are communicated over digital channels, often at different transmission rates, and therefore a signal must be transformed or *coded* for efficiency and compatibility. *Source coding* is the term used for this process.

A *coding system* is a mapping that specifies how the *encoder* transforms source symbols or groups of such signals into a new set of symbols. The *decoder* processes the coded sources to retrieve the initial information. The term "code" is used by information theorists and is usually associated with a lookup table or a dictionary listing the pairs of input/output words or symbols [94, 40, 44].

1.4.1 Entropy coding

In a *fixed rate* coder a constant number of channel bits per time unit is produced by the encoder and processed by the decoder. In certain communication systems fixed rate operation is not satisfactory if the data source has wide variations in activity. For example sampled speech at fixed rate is redundant in moments of silence. A variable rate code is a solution to efficient encoding. An *entropy coding* technique is an efficient variable rate encoder that takes advantages of statistical redundancies in the source whether by prior knowledge or by building a model as it scans the source. Discrete data is encoded into variable length codewords in an invertible manner. The entropy coding is also called *noiseless coding*, *lossless coding*, or *data compaction coding*. The term *data compression* means entropy coding. A typical example of entropy

23

coding is the Morse code: letters with higher occurrence (probability) have corresponding shorter codewords than less probable letters.

Let X be a source whose symbols occur with probability p_k. The *entropy* of this source is defined as

$$\mathcal{H}(X) = - \sum_{1 \leq k \leq M} p_k \log_2 p_k$$

and gives a lower bound on the average bit rate required to code the source. This result is a theorem of Shannon who set the foundations of information theory [40]. However, the information conveyed by Shannon's theorem does not lead to an algorithm for efficient coding [128, 40, 44, 56].

Figure 1.6 shows a high level structure of a transform coder. A trans-
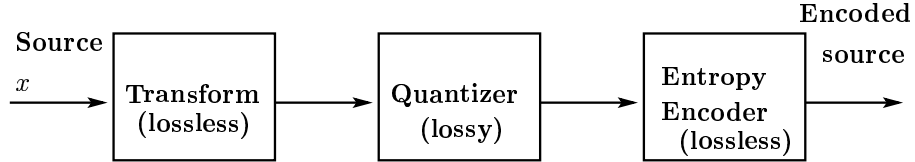
Figure 1.6: High level structure of a transform encoder

form maps the source signal into a more compact representation. This process is usually a decorrelation. The transform is invertible in the sense that in the absence of a quantizer the coded source can be reconstructed without loss of information. The quantizer maps the transformed signal into a smaller set of code words so that the entropy encoder is more efficient. The distortion that arises in a compression process is due solely to the quantization process. The JPEG-2000 Standard generates lossless compression in the presence of a transform that maps integers to integers [24, 109].

In image processing, lossy coding is the most commonly encountered type of compression [5, 14, 20, 44]. Lossless coding is desired by every user of a coding system, although it doesn't always render a satisfactory compression ratio. The perceptually lossless coding, i.e. lossy coding that seems lossless to human auditory or visual systems, is a good trade-off. See [20, 129, 55, 5] for examples of lossy coding strategies that are perceptually lossless.

1.4.2 Quantization theory

Coding processes that involve distortion or loss of information are modeled by a quantizer. The translation of physical experiments into computational systems, like the conversion from analog to digital, suffers from quantization.

A quantizer (Figure 1.7) can be viewed as the composition of two maps: an encoder and a decoder

$$\mathcal{E} : \mathbb{R} \to \mathcal{I} \subset \mathbb{N}$$

$$\mathcal{D} : \mathcal{I} \to \mathcal{C} \subset \mathbb{Q}$$

$$Q = \mathcal{D} \circ \mathcal{E}$$

The *encoder* $\mathcal{E}$ assigns to each real number an index (bin number) in $\mathcal{I}$ that indexes the bin containing x. The decoder $\mathcal{D}$ takes the bin index and returns the quantized value which is some fixed value from each bin:

$$Q(x) = \mathcal{D}(\mathcal{E}(x)) = \mathcal{D}(k) = y$$

25
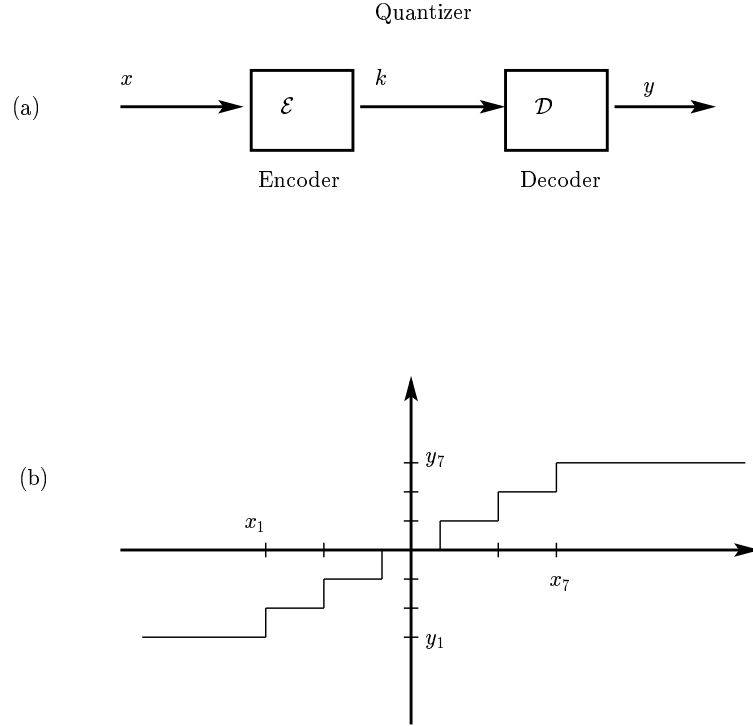
Figure 1.7: (a) Block diagram for quantization; (b) the step nonlinear behaviour of a simple uniform quantizer.

Unlike linear systems, for which there exists a well established theory, such understanding is unavailable for even simple nonlinear quantization processes [40].

We now introduce concepts pertaining to the performance of a quantizer. The most commonly encountered distortion measure between a number x and an approximation of it, $\widehat{x} = Q(x)$, is

$$d(x, \hat{x}) = |x - \hat{x}|^2 \,.$$

26

A more general expression for the distortion is

$$d_m(x, \hat{x}) = |x - \hat{x}|^m$$

for any $m > 0$. The cases when m is 2 or 1 define the squared error or the absolute error, respectively. Most informative measures take into account statistics such as the mean-square error, or the average distortion

$$d = E\{[x - Q(x)]^2\} = \int_{\mathbb{R}} [x - Q(x)]^2 f_X(x) dx$$

where $f_X(x)$ is the probability density function of x. A widely used measure of distortion is the signal-to-noise ratio (SNR) measured in decibels (dB) and defined as

$$\mathrm{SNR} = 10 \log_{10} \frac{E\{x^2\}}{E\{[x - Q(x)]^2\}} .$$

1.4.3   Rate allocation

Next we state the problem of optimal allocation and show how interdependent the distortion measure and the quantization are. Consider a set of M random variables with zero mean and variances $\sigma_1^2, \sigma_2^2, \ldots, \sigma_M^2$, and with known probability density function. Suppose we seek to quantize these variables, given a particular distortion measure and a target average number of bits per sample – also called "bit budget". The problem of optimal bit allocation is the following: find the bit budgets (or rates) $b_1, b_2, \ldots, b_M$ that minimize the overall distortion as a function of the bit allocation vector $\boldsymbol{b} = [b_1, b_1, \ldots, b_M]^T$, subject to the constraint that the average rate is limited

27

to a given quota: $\sum_{k=1}^{M} b_k \leq B$. Given a model for distortion as a function of rates, the solution of the optimal bit allocation problem is rendered by the Lagrange multipliers strategy [109, 40, 100]. In this way, the distortion measure influences the quantization strategy.

We can define a discrete stochastic process $q : \mathbb{Z} \to \mathbb{C}$ by

$$q[n] = Q(x[n]) - x[n]$$

called the *quantization error* or the *noise source* associated with the quantizer. High resolution quantization theory addresses the issues related to the modeling of quantization distortion. High resolution – or high rate – quantization is understood to mean that the average distortion is lower than the input variance, σ_x^2

$$d \ll \sigma_x^2 = E\{(x - \mu_x)^2\}\,.$$

A value of the signal-to-noise ratio of 10 dB is considered the borderline between low and high resolution [40].

1.4.4 High rate quantization theory

We next describe the model of a high resolution quantizer. Assume that the input signal is such that within each quantizer bin the probability density function is approximately constant. In other words, we assume that the quantization bins are fine enough that the input probability density function is modeled reasonably well by a step function defined in the quantization bins. Denote as in the previous section by σ_x^2 the input signal variance of a signal x that is input to a quantizer Q.

28

The variance of the quantizer noise is expressed in this case by the *standard noise model*

$$\sigma_q^2 = c2^{-2b}\sigma_x^2$$

where b is the number of bits at which the channel operates and c is a constant associated with the probability density function of the input. The standard noise model is used very widely to estimate distortion [40, 117].

1.5 Coding schemes

1.5.1 Transform coding

Suppose a block of M consecutive samples of a stationary random process or a vector random variable are subject to coding and a fixed number of bits is specified. In general this signal has correlated components, which implies that the representation of the data is inefficient [40]. The signal can be pre-processed so that the encoder receives uncorrelated components. Suppose a transform acts on the input signal as

$$\boldsymbol{y}[n] = \boldsymbol{T}\boldsymbol{x}[n] \tag{1.12}$$

The bit allocation problem is now solved for the transformed signal $\boldsymbol{y}$. The quantizers are $Q_0, Q_1, \ldots, Q_{M-1}$. Such a scheme is called a *transform coder* (see Figure 1.8). The transform matrix $\boldsymbol{T} \in \mathbb{C}^{M \times M}$ is usually unitary. In most practical applications the transform is real orthogonal.

The performance of a transform coder is expressed by the *coding gain* (see [40, 117]). The coding gain of a transform coder under the high-bit rate
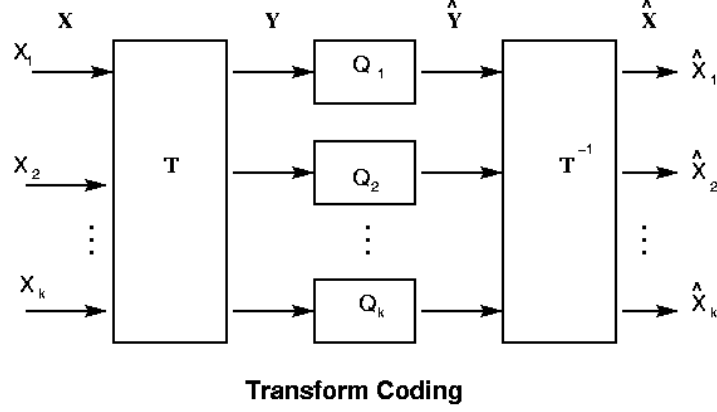
Transform Coding

Figure 1.8: Transform coder/decoder

assumption is given by the relation

$$G_{TC} = \frac{\sigma_x^2}{(\prod_{k=1}^{M} \lambda_{y_k}^2)^{1/M}} \tag{1.13}$$

where σ_x^2 is the input variance and $\lambda_{y_k}^2, k = 1, 2, \ldots, M$ are the eigenvalues of the auto-covariance matrix of the transformed signal $\boldsymbol{y}$. An optimal transform coder is presented in the next chapter. The performance of a transform coder depends on the choice of the transform for a given input signal. For details see [72, 71, 125].

1.5.2 Sub-band coding

A strategy that is similar to the transform coder but uses filters with memory is called sub-band coding. A sub-band coding scheme is shown in Figure 1.9. A signal is decomposed or analyzed in components that are tailored to the specific quantization problem for which the sub-band coder is sought to be optimal. The components of the signal are called *sub-bands*. After analysis,

30
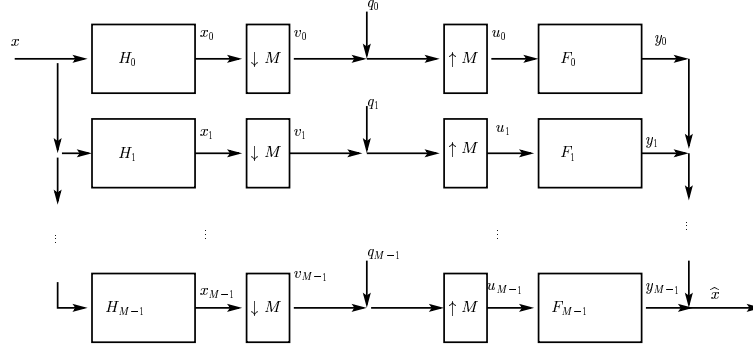
Figure 1.9: Sub-band coder

a quantization takes place; reconstruction of the signal is accomplished by the synthesis filter bank. Filter banks and sub-band coders are viewed in the digital processing communities as closely related. And that is because a perfect reconstruction filter bank without a quantizer makes little sense in practice. However, there are applications where the filter banks do not have the perfect reconstruction property – it depends on the practical problem [120]. We restrict our discussion to perfect reconstruction filter banks.

There are advantages of the sub-band coding strategy over the transform coder. Detailed discussion can be found in [125, 61, 10, 127]. A sub-band coder is characterized by its coding gain, defined as the coding gain of the transform coder [117]. In general, a sub-band coder renders a higher coding gain than a transform coder. The coding gain of a sub-band coder is

$$G_{SBC} = \frac{(1/M)\sum_{k=0}^{M-1}\sigma_{v_k}^2}{(\prod_{k=0}^{M-1}\sigma_{v_k}^2)^{1/M}} \tag{1.14}$$

where $\sigma_{v_k}^2$, $k = 0, 1, \ldots, M - 1$, are the sub-band variances from the analysis bank. The expression for the coding gain of a sub-band coder is defined as the

31

ratio

$$G_{SBC} = \frac{\epsilon_{direct}}{\epsilon_{SBC}} \qquad (1.15)$$

where ϵ_{SBC} is the mean-square value of the reconstruction error $\widehat{x}(\cdot) - x(\cdot)$ in the presence of a transform or filter bank and ϵ_{direct} is the mean-square value of the direct quantization error (roundoff quantization without any transform applied prior to quantization), with the same average bit rate b. However, regardless of the quantization process, the coding gain of a filter bank is also defined by Equation (1.14). The transform coder can be viewed as a particular case of a sub-band coder and that justifies the term "sub-band signals" used for the output of a transform in a transform coding scheme.

1.6 Remarks

Above we introduced the essentials of digital signal processing. Our main intention is to familiarize the reader with the notions of filter banks, quantization, sub-band coding, etc. We do not introduce in-depth image compression strategies, although, the tools presented are intended to assist efficient compression of large images. The actual standard compression paradigm is presented by Taubman and Marcellin in [109] and for coding theory see [94, 40]. For further background material in digital signal processing and filter bank theory see [85, 117, 106, 54, 127, 72, 56, 29].

1.7 Thesis Outline

Here we summarize the content of the remaining chapters.

Chapter 2: The Karhunen-Loève Transform

The use of the Karhunen-Loève transform dates back to the first digital signal processing problems. We use this transform extensively to decorrelate multicomponent images prior to compression. For hyper-spectral images, the Karhunen-Loève transform is the best decorrelator [92, 54, 59, 21]. However, the Karhunen-Loève transform is not suited in transform coder schemes with non-Gaussian input such as medical three-dimensional volumetric images. This discovery motivated a deeper study of the transform. We describe the continuous and discrete Karhunen-Loève transform, and optimality and computational issues.

Chapter 3: Filter Banks and Optimality Issues

This chapter is intended to familiarize the reader with the filter bank theory and the usual criteria of optimality of filter banks used by the digital signal processing community [118]. Any sequence of image processing tasks that starts with a Karhunen-Loève transform is, in general, subject to an upgrade using a filter bank. This background is essential to the understanding of the existence and design of a filter bank that exhibits properties similar to the Karhunen-Loève transform with respect to decorrelation. However, there is an optimality criterion, the *coding gain*, that is very widely used; it expresses the ability of a filter bank to render a signal that is susceptible to efficient compression, similar to the principal component criterion. This chapter focuses on optimality in the coding gain sense [63, 119]. The classes of filters studied are limited to perfect reconstruction orthonormal filter banks. The theory

33

developed by Vaidyanathan [118] pertains to unrealizable ideal filter banks; however it is valuable in the performance analysis of realizable filter banks.

Chapter 4: Principal Component Filter Banks

We begin with the mathematical setting for optimality study from the principal component perspective. The results from the previous chapters on the Karhunen-Loève transform and filter bank theory are now woven into a more powerful optimality concept, in the principal component sense. While there are principal component filter banks in the class of unrealizable filter banks, there is no definitive answer to the question of existence of these filters in the class of realizable filter banks. We discuss the existence of principal component filter banks based on the results in the literature and our experiments.

In a survey paper Akkarakaran and Vaidyanathan [10] state that there are no coding gain optimal filter banks for realizable filters. We study the existence of principal component filter banks for a model signal and realizable filter banks in response to both [10] and a study case that was incorrectly set by Kirac and Vaidyanathan in [62]. We prove that in any class of realizable filter banks there exist filter banks that are optimal with respect to the coding gain. We also present an algorithm to find these optimal filter banks analytically. These are our main contributions to the theory and design of optimal filter banks. On the practical side, our results address the issue of existence and design of principal component filter banks for simple structures with only two filters, or two channels. These are the filter banks supported by JPEG-2000 Standard, hence our filter bank design has a broad range of applicability.

Chapter 5: Applications of Optimal Filter Banks to Hyper-spectral Imagery Processing

The compression of hyper-spectral images benefits from the use of the Karhunen-Loève transform as a decorrelator in the spectral direction. We show in [86] that there are applications from the class of pattern recognition problems where data exploitation performed on compressed data is almost as good as tasks performed on uncompressed data. A hyper-spectral image with a size of 150MB was compressed with a ratio of 128 resulting in a compressed image with a size less than 1 percent of the initial data. Some widely used feature extractions matched within 99.5% the features retrieved from the uncompressed data. The applicability of feature extraction from highly compressed images is primary in real-time applications when transmission time and bandwidth are essential for good information exchange.

Principal component filter banks always exist for two-channel filter banks. Once the power spectral density of the input is known, an optimal filter bank can be found analytically using symbolic computation packages. We design and compare the performances of two-channel principal component filter banks and the Karhunen-Loève transform for a class of hyper-spectral images.

Chapter 2

The Karhunen-Loève Transform

2.1   Motivation

This transform was introduced originally by Karhunen [58] and Loève [68] as a series expansion for continuous random processes. The optimality of this transform in the mean square sense is exhibited both for continuous and discrete time signals.

For random sequences, Hotelling [49, 50] discusses what was first called a method of principal components, which in essence is a discrete version of the Karhunen-Loève series expansion; that is why the Karhunen-Loève transform is also called the Hotelling transform, or principal component transform in many scientific communities [92, 17, 99]. The principal component transform – or principal component analysis – is also closely related to other decorrelation techniques, such as independent component analysis [64, 52].

In a coding scheme, the Karhunen-Loève transform is optimal with respect to the coding gain defined in Equation (1.13). In remote sensing applications this transform is used to analyze images in a feature space that is free of redundancies; it can also be used for color contrast enhancement. Among the many applications of the Karhunen-Loève transform in computer

vision we mention pattern recognition, object tracking and motion estimation.

The Karhunen-Loève transform is very widely used; however it poses two major challenges to users. First, it is a signal dependent transform since it depends on the auto-covariance matrix; thus the basis has to be computed separately for every input. Second, it is a very expensive transform: no structure can be assumed for the covariance matrix other than being symmetric and positive definite. Therefore no fast algorithm can be used; the cost of constructing the Karhunen-Loève transform is of order $O(N^3)$.

We present the continuous and the discrete versions of the transform and its use in signal coding. We also outline alternate strategies to overcome the high cost. Theoretical issues are discussed further in Chapter 4 from the perspective of filter bank optimality.

2.2   The Continuous-Time Karhunen-Loève Transform

We follow the literature and limit the presentation to the real case [75]. For the complex case all the results are similar. The continuity concept used in this section is defined next.

Definition 2.2.1. The random process x is continuous at t_0 in the mean-square sense if

$$\text{l.i.m.}_{t \to t_0} x(t) = x(t_0)$$

where l.i.m. denotes the *limit in the mean*:

$$\lim_{t \to t_0} E\{[x(t) - x(t_0)]^2\} = 0 \,. \tag{2.1}$$

37

Consider a real-valued continuous-time stochastic process $x : [a, b] \to \mathbb{R}$ with a finite mean-square value (or finite variance) $E\{x^2\} < \infty$ and with zero mean $E\{x(t)\} = 0$ for all $t \in [a, b]$. We seek a series expansion that represents the stochastic process in the mean as in Definition 2.2.1

$$x(t) = \text{l.i.m.}_{M \to \infty} \sum_{k=1}^{M} x_k \psi_k(t) \tag{2.2}$$

or, equivalently, from Definition 2.2.1

$$\lim_{M \to \infty} E\{[x(t) - \sum_{k=1}^{M} x_k \psi_k(t)]^2\} = 0, \qquad t \in [a, b]. \tag{2.3}$$

The orthonormal basis $\psi_k(t); k = 1, 2, \ldots$ in Equation (2.1) will be determined from the properties of the stochastic process as follows. The coefficients x_k will be the Fourier coefficients

$$x_k = \langle x, \psi_k \rangle = \int_a^b x(t) \psi_k(t) dt \,.$$

Let us assume that $E\{x_k\} = 0$, $k = 1, 2, \ldots$ We require that the coefficients of the expansion be uncorrelated:

$$
\begin{aligned}
\lambda_l \delta_{kl} &= E\{x_k x_l\} \tag{2.4} \\
&= E\{\langle x, \psi_k \rangle \langle x, \psi_l \rangle\} \\
&= E\left\{ \left(\int_a^b x(t) \psi_k(t) dt \right) \left(\int_a^b x(u) \psi_l(u) du \right) \right\} \\
&= E\left\{ \int_a^b \psi_k(t) \int_a^b x(t) x(u) \psi_l(u) du \, dt \right\} \\
&= \int_a^b \psi_k(t) \left(\int_a^b E\{x(t) x(u)\} \psi_l(u) du \right) dt
\end{aligned}
$$

38

The kernel of the integral operator in Equations (2.4) is the symmetric auto-correlation function

$$R_{xx}(t, u) = E\{x(t)x(u)\} \tag{2.5}$$

and we rewrite the condition in Equation (2.4) as

$$\lambda_l \delta_{kl} = \int_a^b \psi_k(t) \left(\int_a^b R_{xx}(t, u)\psi_l(u)du \right) dt \,. \tag{2.6}$$

Equation (2.6) holds if

$$\int_a^b R_{xx}(t, u)\psi_l(u)du = \lambda_l \psi_l(t) \,. \tag{2.7}$$

The solutions $\psi_l(t), \quad l = 1, 2, \ldots$ of the integral Equation (2.7) form the orthonormal basis that guarantees that the coefficients of the expansion are uncorrelated. The integral operator in Equation (2.7) has eigenfunctions $\psi_l(t), \ l = 1, 2, \ldots$ with the associated eigenvalues $\lambda_l, \ l = 1, 2, \ldots$.

The kernel $R_{xx}(\cdot, \cdot)$ is a covariance function and therefore it is symmetric and nonnegative definite. With the hypothesis that the variance of the input is finite, the covariance function satisfies [37, 112]

$$\int_a^b \int_a^b R_{xx}^2(t, u)dtdu \leq \left(\int_a^b E\{x^2(t)\}dt \right)^2 < \infty$$

We now invoke results from linear operator theory [104, 93].

- There exist at least one square integrable function $\psi(\cdot)$ and a real number λ that satisfy Equation (2.7).

- if $\psi(\cdot)$ is a solution of Equation (2.7) then $c\psi(\cdot)$ is also a solution and the eigenfunctions can thus be normalized.

- Eigenfunctions corresponding to distinct eigenvalues are orthogonal to each other.

- The set of eigenfunctions is countable and they all are bounded almost everywhere.

- If $\psi_1(\cdot)$ and $\psi_2(\cdot)$ are eigenfunctions associated with the same eigenvalue λ then $c_1\psi_1(\cdot) + c_2\psi_2(\cdot)$ is also an eigenfunction associated with λ.

- For any eigenvalue λ there is at most a finite number of linearly independent eigenfunctions.

- If the process x is continuous at each $t \in [a, b]$ in the mean-square sense, then the kernel $R_{xx}(\cdot, \cdot)$ is continuous on $[a, b] \times [a, b]$ – see [88]. Mercer's Theorem (see Riesz and Sz.-Nagy [93] and Brislawn [18]) states that the kernel $R_{xx}(\cdot, \cdot)$ can be expanded in the series

$$R_{xx}(t, u) = \sum_{k=1}^{\infty} \lambda_k \psi_k(t) \psi_k(u), \qquad a \leq t, u \leq b$$

and the convergence of the series is uniform for all t and u in $[a, b]$.

- The sum of the eigenvalues is the expected value of the energy of the zero mean process $x(\cdot)$ in the interval $[a, b]$

$$E\left\{ \int_a^b x^2(t)dt \right\} = \int_a^b R_{xx}(t, t)dt = \sum_{k=1}^{\infty} \lambda_k \, .$$

Now we show that the convergence in the mean from Equation 2.3 holds. Denote by $\epsilon_M(\cdot)$ the expected value of the mean square error when $x(\cdot)$ is

40

approximated by the first M terms

$$\epsilon_M(t) = E\left\{[x(t) - \sum_{k=1}^{M} x_k\psi_k(t)]^2\right\}$$

Evaluate

$$
\begin{aligned}
\epsilon_M(t) &= R_{xx}(t,t) - 2E\left\{x(t)\sum_{k=1}^{M} x_k\psi_k(t)\right\} + E\left\{\sum_{k=1}^{M}\sum_{l=1}^{M} x_k x_l\psi_k(t)\psi_l(t)\right\} \\
&= R_{xx}(t,t) - 2E\left\{x(t)\sum_{k=1}^{M}(\int_a^b x(u)\psi_k(u)du)\psi_k(t)\right\} \\
&\quad + \sum_{k=1}^{M}\lambda_k\psi_k(t)\psi_k(t) \\
&= R_{xx}(t,t) - 2\sum_{k=1}^{M}(\int_a^b R_{xx}(t,u)\psi_k(u)du)\psi_k(t) + \sum_{k=1}^{M}\lambda_k\psi_k(t)\psi_k(t) \\
&= R_{xx}(t,t) - \sum_{k=1}^{M}\lambda_k\psi_k(t)\psi_k(t)
\end{aligned}
$$

Mercer's Theorem guarantees that the last sum converges uniformly to $R_{xx}(t,t)$ for t in $[a,b]$ as M approaches infinity. We proved that

$$\lim_{M\to\infty}\epsilon_M(t) = 0, \qquad \text{uniformly for } t \in [a,b].$$

In practice, solving the integral equation to determine the basis is a difficult task; therefore the continuous Karhunen-Loève transform is only valuable as a theoretical tool.

We can evaluate only M terms in Equation (2.2) and obtain an approximation of the signal to be represented. The approximation error has mean equal to the sum of the eigenvalues not used in the approximate representation

41

– and that sum is finite – therefore we consider the largest eigenvalues for a finite representation.

2.3   The Discrete Karhunen-Loève Transform

In digital signal processing the discrete version of the Karhunen-Loève transform is used [75, 72, 54, 125, 88].

Consider a real-valued M-dimensional vector random process.

$$\boldsymbol{x} = [x_1 \ ... \ x_M]^T, \qquad \boldsymbol{x} \in \mathbb{R}^M .$$

We seek an orthogonal transform that is optimal in the mean-square error criterion. Let $\boldsymbol{U}$ be an orthogonal transform given by

$$\boldsymbol{U}^T = [\boldsymbol{\varphi}_1 \ \boldsymbol{\varphi}_2 \ ... \ \boldsymbol{\varphi}_M]$$

where the $\boldsymbol{\varphi}_k$ are vectors in $\mathbb{R}^M$ that form an orthonormal basis: $\langle \boldsymbol{\varphi}_k, \boldsymbol{\varphi}_l \rangle = \delta_{kl}$ or $\boldsymbol{U}^T \boldsymbol{U} = \boldsymbol{U} \boldsymbol{U}^T = \boldsymbol{I}$. Let $\boldsymbol{y} \in \mathbb{R}^M$ be the transform of $\boldsymbol{x}$

$$\boldsymbol{y} = \boldsymbol{U} \boldsymbol{x} .$$

Equivalently

$$\boldsymbol{x} = \boldsymbol{U}^T \boldsymbol{y} = [\boldsymbol{\varphi}_1 \ \boldsymbol{\varphi}_2 \ ... \ \boldsymbol{\varphi}_M] \boldsymbol{y}$$

or

$$\boldsymbol{x} = \sum_{k=1}^{M} y_k \boldsymbol{\varphi}_k . \tag{2.8}$$

The vector $\boldsymbol{y}$ can be regarded as the (vector of) coordinates of $\boldsymbol{x}$ in the basis $\boldsymbol{U}^T$. We retain a subset $\{y_1, y_2, ..., y_P\}_{P \leq M}$ of the components of $\boldsymbol{y}$ and

estimate $\boldsymbol{x}$. For the missing $M - P$ components of $\boldsymbol{y}$ we seek constants b_k, $P + 1 \le k \le M$ and obtain

$$\widehat{\boldsymbol{x}}(P) = \sum_{k=1}^{P} y_k \boldsymbol{\varphi}_k + \sum_{k=P+1}^{M} b_k \boldsymbol{\varphi}_k$$

where $\widehat{\boldsymbol{x}}(P)$ is the estimate of $\boldsymbol{x}$. The error vector $\Delta \boldsymbol{x}$ induced by replacing the components y_k with b_k, $k = P + 1, ..., M$, is

$$\Delta \boldsymbol{x} = \boldsymbol{x} - \widehat{\boldsymbol{x}}(P) = \sum_{k=P+1}^{M} (y_k - b_k) \boldsymbol{\varphi}_k \,.$$

The mean-square error is

$$
\begin{aligned}
\epsilon(P) &= E\{\|\Delta \boldsymbol{x}\|^2\} = E\{(\Delta \boldsymbol{x})^T (\Delta \boldsymbol{x})\}, \qquad \text{or, equivalently,} \\
\epsilon(P) &= E\left\{ \sum_{k=P+1}^{M} \sum_{l=P+1}^{M} (y_k - b_k)(y_l - b_l) \boldsymbol{\varphi}_k^T \boldsymbol{\varphi}_l \right\},
\end{aligned}
$$

which simplifies to

$$\epsilon(P) = \sum_{k=P+1}^{M} E\{(y_k - b_k)^2\} \,. \tag{2.9}$$

We now seek an optimal choice of $\boldsymbol{\varphi}_k$ and b_k that minimizes $\epsilon(P)$ for all $P = 1, 2, \ldots, M$. This is optimality in the mean-square sense. Another term used is optimal mean-square projection error sense. These optimal entities are found in two steps:

Step 1. The optimum b_k's are determined from the condition

$$\frac{\partial}{\partial b_k} E\{(y_k - b_k)^2\} = 0 \,.$$

From

$$\frac{\partial}{\partial b_k} E\{(y_k - b_k)^2\} = -2(E\{y_k\} - b_k)$$

43

we obtain the b_k's as

$$b_k = E\{y_k\}.$$

From $y_k = \boldsymbol{\varphi}_k^T \boldsymbol{x}$ we have $b_k = \boldsymbol{\varphi}_k^T E\{\boldsymbol{x}\}$ or

$$b_k = \boldsymbol{\varphi}_k^T \boldsymbol{\mu}, \qquad k = 1, 2, \ldots, M$$

where $\boldsymbol{\mu}$ denotes the expectation of $\boldsymbol{x}$, i.e., $\boldsymbol{\mu} = E\{\boldsymbol{x}\}$. The quantities $y_k - b_k$ are scalars and we can write alternatively in the form

$$
\begin{aligned}
\epsilon(P) &= \sum_{k=P+1}^{M} E\{(y_k - b_k)(y_k - b_k)^T\} \\
&= \sum_{k=P+1}^{M} \boldsymbol{\varphi}_k^T E\{(\boldsymbol{x} - \boldsymbol{\mu})(\boldsymbol{x} - \boldsymbol{\mu})^T\} \boldsymbol{\varphi}_k.
\end{aligned}
$$

Denote by $\boldsymbol{C_{xx}}$ the auto-covariance matrix of $\boldsymbol{x}$

$$\boldsymbol{C_{xx}} = E\{(\boldsymbol{x} - \boldsymbol{\mu})(\boldsymbol{x} - \boldsymbol{\mu})^T\}.$$

and write the error as

$$\epsilon(P) = \sum_{k=P+1}^{M} \boldsymbol{\varphi}_k^T \boldsymbol{C_{xx}} \boldsymbol{\varphi}_k.$$

Step 2. The optimum $\boldsymbol{\varphi}_k$ must minimize $\epsilon(P)$ and obey the orthonormality constraint $\boldsymbol{\varphi}_k^T \boldsymbol{\varphi}_l = \delta_{kl}$. We use the method of Lagrange multipliers: minimize

$$
\begin{aligned}
\hat{\epsilon}(P) &= \epsilon(P) - \sum_{k=P+1}^{M} \lambda_k [\boldsymbol{\varphi}_k^T \boldsymbol{\varphi}_k - 1] \\
&= \sum_{k=P+1}^{M} [\boldsymbol{\varphi}_k^T \boldsymbol{C_{xx}} \boldsymbol{\varphi}_k - \lambda_k(\boldsymbol{\varphi}_k^T \boldsymbol{\varphi}_k - 1)]
\end{aligned}
$$

44

with respect to $\boldsymbol{\varphi}_k$ where the λ_k's are the Lagrange multipliers. We have

$$\nabla_{\boldsymbol{\varphi}_k}[\boldsymbol{\varphi}_k^T \boldsymbol{C_{xx}}\boldsymbol{\varphi}_k] = 2\boldsymbol{C_{xx}}\boldsymbol{\varphi}_k \qquad \text{and}$$

$$\nabla_{\boldsymbol{\varphi}_k}[\boldsymbol{\varphi}_k^T \boldsymbol{\varphi}_k] = 2\boldsymbol{\varphi}_k$$

and thus

$$\nabla_{\boldsymbol{\varphi}_k}[\hat{\epsilon}(P)] = 2\boldsymbol{C_{xx}}\boldsymbol{\varphi}_k - 2\lambda_k\boldsymbol{\varphi}_k$$

or equivalently

$$\boldsymbol{C_{xx}}\boldsymbol{\varphi}_k = \lambda_k\boldsymbol{\varphi}_k, \qquad k = 1, ..., M. \tag{2.10}$$

The equation above tells us that each $\boldsymbol{\varphi}_k$ is the eigenvector of the covariance matrix $\boldsymbol{C_{xx}}$ and λ_k is the corresponding eigenvalue. Finally, when the eigenvalues are ordered decreasingly, the minimum mean-square error, $\epsilon_{min}(\cdot)$, is

$$\epsilon_{min}(P) = \sum_{k=P+1}^{M} \lambda_k. \tag{2.11}$$

The expansion defined in Equation (2.8) is called the Karhunen-Loève expansion.

This eigenvalue problem has the following properties:

1. $\boldsymbol{C_{xx}}$ is a covariance matrix, therefore it is positive semi-definite; hence the eigenvalues λ_k are real and non-negative.

2. Eigenvectors that belong to distinct eigenvalues are orthogonal to each other.

3. Multiple eigenvalues have linearly independent eigenvectors and they can be chosen to be orthogonal to each other.

The normalized eigenvectors comprise the basis of the Karhunen-Loève transform. For the complex case the covariance matrix:

$$\boldsymbol{C_{xx}} = E\{\boldsymbol{xx}^\dagger\}$$

is also symmetric and non-negative definite [38, 37]. The Karhunen-Loève transform matrix $\boldsymbol{U}$ is unitary and satisfies $\boldsymbol{UU}^\dagger = \boldsymbol{I}$. We cannot conclude that the imaginary and real part are uncorrelated; we can only say that

$$E\{y_k y_l^*\} = \lambda_k \delta_{kl}, \qquad k, l = 1, 2, \ldots, M.$$

The covariance matrices of the input signal and the transformed signal are related as

$$\boldsymbol{U C_{xx} U}^\dagger = \boldsymbol{\Lambda}\,,$$

$$\boldsymbol{C_{xx}} = \boldsymbol{U}^\dagger \boldsymbol{\Lambda U}\,.$$

The matrix $\boldsymbol{\Lambda}$ is the diagonal matrix of eigenvalues of the covariance matrix.

$$\boldsymbol{\Lambda} = \text{diag}\,\{\lambda_1,\ \lambda_2,\ \ldots,\ \lambda_M\}\,.$$

If the input vector process $\boldsymbol{x}$ has zero mean (see Section 1.2.2) the Karhunen-Loève transform diagonalizes the autocorrelation matrix of the process.

2.4   Properties of the Karhunen-Loève Transform and Optimality Issues

A similar approach as for the continuous case can be considered for the discrete case. We may define the discrete Karhunen-Loève transform by

requiring that the components of the output vector process be uncorrelated. The output auto-covariance matrix is then diagonal and we conclude that the transform we seek must diagonalize the input auto-covariance matrix. We then show that an approximation in such a basis is optimal in the mean-square sense using the same Lagrange multipliers technique . This is the approach presented by Mertins [75] and Jain [54].

Brillinger [17] opens a new approach to signal processing optimization techniques. The aim is to approximate a series by a filtered version of itself with a filter having reduced rank. This is the principal component approach for time series analysis. For transform coding schemes such an approach is an optimization in the mean square sense for a basis restriction [54].

2.4.1 Distribution of variances and the minimal geometric mean property

Proposition 2.4.1. *Let a Gaussian random vector process be input to unitary $M \times M$ transforms and assume that the variances of the output transformed signal are ordered decreasingly. Then the Karhunen-Loève transform outputs the largest first P variances, for $P = 1, 2, \ldots, M$.*

Proof. Let $\boldsymbol{T}$ be an arbitrary unitary transform and $\boldsymbol{U}$ be the Karhunen-Loève transform of an input M-dimensional zero mean Gaussian discrete random vector $\boldsymbol{x}$. Let $\boldsymbol{y}$ be the output from the transform $\boldsymbol{T}$,

$$\boldsymbol{y} = \boldsymbol{T}\boldsymbol{x}\,.$$

Define the average variances component-wise and order them decreasingly:

$$\sigma_k^2 = E\{|y_k|^2\}, \qquad \sigma_1^2 \geq \sigma_2^2 ... \geq \sigma_M^2 .$$

Let P be in $\{1, 2, \ldots M\}$ and denote the partial sum of output variances (or partial energy)

$$S_P(\boldsymbol{T}) = \sum_{k=1}^{P} \sigma_k^2 .$$

We write the partial energy in terms of the input and the arbitrary transform

$$
\begin{aligned}
S_P(\boldsymbol{T}) &= \sum_{k=1}^{P} E\{[(\boldsymbol{Tx})(\boldsymbol{Tx})^\dagger]_{k,k}\} \\
&= \sum_{k=1}^{P} [\boldsymbol{T}\boldsymbol{C_{xx}}\boldsymbol{T}^\dagger]_{k,k} \\
&= \operatorname{tr}(\boldsymbol{I}_P\boldsymbol{T}^\dagger\boldsymbol{C_{xx}}\boldsymbol{T})
\end{aligned}
$$

where

$$\boldsymbol{I}_p = \operatorname{diag}\{1, 1, \ldots, 1, 0, 0, \ldots, 0\}$$

is obtained by bordering with zero the identity matrix of order P. By the results of Section 2.3, the sum $S_P(\boldsymbol{T})$ is maximized when $\boldsymbol{T}$ is the Karhunen-Loève transform of the input. $\qquad\square$

Next we show the minimal geometric mean property of the Karhunen-Loève transform: for any positive semidefinite matrix $\boldsymbol{C} \in \mathbb{R}^{M \times M}$ we have (see [48])

$$\det \boldsymbol{C} \leq \prod_{k=1}^{M} c_{kk}$$

48

where the c_{kk}'s are the diagonal elements of $\boldsymbol{C}$. Equality holds when $\boldsymbol{C}$ is diagonal. Since the Karhunen-Loève transform leads to a diagonal covariance matrix of the representation, we conclude that in fact Karhunen-Loève transform leads to random processes with a minimal geometric mean of the variances. From Equation 1.13, we therefore conclude that the Karhunen-Loève is the optimal transform for transform coding schemes [54, 40, 56, 117].

2.4.2 Uniqueness of the Karhunen-Loève transform

The Karhunen-Loève basis is not unique if eigenvalues of multiplicity greater than 1 are present. For the special case when $\boldsymbol{Cxx}$ is the covariance matrix of a white noise input process with

$$\boldsymbol{Cxx} = \sigma^2 \boldsymbol{I}\,,$$

then $\lambda_1 = \lambda_2 = ... = \lambda_M = \sigma^2$. A white noise process can be optimally approximated with any orthogonal basis.

2.5 A First Quantization Result for Optimal Sub-band Coding

Huang and Schultheiss [51] show that the Karhunen-Loève transform is the optimal transform coder under mild assumptions on the quantization noise sources. Their study shows that optimization results for coding gain are not guaranteed if the input samples do not have a jointly Gaussian probability distribution. The results in [51] are the first ones to show that the performance of a decorrelation transform is independent of the loss due to quantization.

49
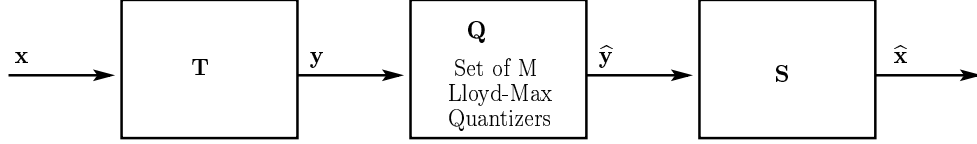
Figure 2.1: Transform coder with Max-Lloyd quantizers.

We refer to Figure 2.1. The assumptions made are the following:

- The transform coder T has as input a signal x modeled as a Gaussian M- dimensional real random variable with zero mean.

- The matrix T that operates on x yields an M-dimensional uncorrelated random variable y. Note that the random variable y is also Gaussian because the input is Gaussian and the transform T is linear. In fact, the random variable y has components that are not only uncorrelated but also independent [88, 52].

- The sub-band processors P_k are b_k-bit Lloyd-Max quantizers [40] and they are optimal in the sense that the quantized sub-band $\hat{y}_k$ is orthogonal to the quantization error $\epsilon_k = y_k - \hat{y}_k$; that is $E\{\epsilon_k \hat{y}_k\} = 0$. No high bit-rate is assumed.

- The quantizer Q is such that the quantization error is minimized subject to an overall fixed bit budget [40, 67, 66, 74].

- Sub-bands are numbered such that the variances are ordered decreasingly and so are the bit budgets allowed for each sub-band signal quantization.

The Karhunen-Loève transform enters in the area of transform coding as the solution to an optimization problem. Given a fixed bit budget b the problem is to find appropriate matrices $\boldsymbol{T}$ and $\boldsymbol{S}$ and bit budget values $b_k, \; k = 1, 2, ..., M$ for each channel such that the mean-square error $E\{(\boldsymbol{x} - \widehat{\boldsymbol{x}})^T (\boldsymbol{x} - \widehat{\boldsymbol{x}}\}$ is minimized. We list the results from [51].

- The matrices $\boldsymbol{T}$ and $\boldsymbol{S}$ are inverse to each other – this is universally the actual choice because it renders prefect reconstruction in the absence of a quantizer.

- Denote by $\boldsymbol{C_{xx}}$ the autocorrelation matrix of the input and let $\lambda_1, \lambda_2, \ldots, \lambda_M$ be its eigenvalues. The matrix $\boldsymbol{T}$ that minimizes the mean square error is the Karhunen-Loève transform of the input

$$\boldsymbol{T} = \boldsymbol{U}^T$$

   where

$$\boldsymbol{U}^T \boldsymbol{C_{xx}} \boldsymbol{U} = \boldsymbol{\Lambda}$$

   and as in Section 2.3 the eigenvalues λ_k are ordered decreasingly. This choice of the matrix $\boldsymbol{U}$ is optimal for any number of bits b_k allocated on the k-th channel satisfying $b_1 \geq b_2 \geq ... \geq b_M$ with the average bit rate b satisfying: $\sum_{k=1}^{M} b_k = Mb$.

- The best choice for the kth bit budget b_k is given approximately by

$$b_k = b + \frac{1}{2} \log_2 \frac{\lambda_k}{(\prod_{k=1}^{M} \lambda_k)^{1/M}} \; .$$

51

Some of the bit budgets b_k's may be negative or non-integer, and to overcome this difficulty iterative rate allocation methods that truncate negative bit budgets and reallocate bits are used in practical applications.

- A lower bound on the mean square error is

$$\epsilon_{min} = 2^{-2b}(\prod_{k=1}^{M} \lambda_k)^{1/M}$$

When the Karhunen-Loève transform was first introduced, there were no constraints set on the input process. Under a linear transform, it is only a Gaussian input that is mapped into a Gaussian output. The process of selecting a quantizer is influenced by the probability density function of the input to the quantizer. Hence if the input to the transform is not Gaussian, the statistics of the output of the transform – and the input to the quantizer – are often not known. Optimization in the non-Gaussian case appears to be a difficult problem.

The Karhunen-Loève transform only takes into account the second order statistics of the process, hence it is optimal only for Gaussian input in coding schemes. Feng and Effros [39] show that the Karhunen-Loève transform is not the optimal orthogonal transform for coding gain if the input has a uniform distribution. A similar observation can be found in Mallat [72].

Further details of the Karhunen-Loève transform are given in Chapter 4.

2.6 Threshold Representation

The Karhunen-Loève transform also minimizes $E\{P\}$, the *expected number of transform coefficients* required for their energy to meet a specified threshold [12]. Consider the Karhunen-Loève expansion of a zero-mean random process x on the interval $a \leq t \leq b$ as in Equation (2.1)

$$\boldsymbol{x}(t) = \text{l.i.m}_{M \to \infty} \sum_{k=1}^{M} x_k \psi_k(t)$$

with the basis functions $\psi_k, \ k = 1, 2, \ldots$ given by Equation (2.7)

$$\int_a^b R_{xx}(t, u)\psi_l(u)du = \lambda_l \psi_l(t) \,.$$

Consider a finite-dimensional approximation to x,

$$\tilde{x}_M(t) = \sum_{k=1}^{M} x_k \varphi_k(t) \tag{2.12}$$

in which $\varphi_k(t), \ k = 1, 2, \ldots, M$ form a complete orthonormal set and the x_k are the corresponding generalized Fourier coefficients [25] for the representation. Define a random variable R_0 as

$$R_0 = \int_a^b x^2(t)dt$$

and let the random variables $R_M, \ M = 1, 2, \ldots$ be the integral-square truncation error from the expansion in Equation (2.12) and with a zero-mean hypothesis on x it follows that

$$R_M = R_0 - \sum_{k=1}^{M} x_k^2 = \int_a^b [x(t) - \sum_{k=1}^{M} x_k \varphi_k(t)]^2 dt \,.$$

53

For a fixed M, the mean truncation error is minimized if the basis functions $\{\varphi_k\}$ of Equation (2.12) satisfy the Karhunen-Loève basis restriction (see [54]). This solution depends only on the correlation operator $R_{xx}(\cdot, \cdot)$ from Equation (2.5). Rather than fixing M and seeking to minimize $E\{R_M\}$, a threshold level ϵ^2 is fixed and it is required that the integral-square truncation error be less than or equal to the threshold level. Let M be a random variable such that $E\{M\}$ is the value of m that satisfies:

$$R_{m-1} > \epsilon^2; \qquad R_m \leq \epsilon^2$$

A complete orthonormal set $\{\varphi_k\}_k$ is sought such that $E\{M\}$ is minimized. The set $\{\varphi_k\}_k$ depends on the statistics of the process; in particular it is not determined just by the correlation function as for the basis restriction optimization criterion. Since the distribution of a Gaussian process depends only on its correlation function one can conjecture that the Karhunen-Loève expansion is again optimal under the criterion of minimizing $E\{M\}$ for a given threshold ϵ^2, and for Gaussian input. A proof of this conjecture can be found in [12] under the additional assumption that the process $x(\cdot)$ has finite energy:

$$E\{R_0\} = \sum_{k=1}^{\infty} \lambda_k < \infty \,.$$

We show at the end of the chapter a few applications of this property of the Karhunen-Loève transform.

54

2.7 Computational Issues

The diagonalization of an $M \times M$ matrix carries a cost of $O(M^3)$. For large collections of data there is a significant cost for the construction or estimation of the auto-covariance matrix. There are alternate routes to overcome these difficulties. We mention a few of them:

- If the input can be modeled by an autoregressive process, the eigenvectors and eigenvalues of the Karhunen-Loève transform can be determined analytically. See Ray and Driver [91].

- When the autocorrelation matrix is circulant the Karhunen-Loève transform reduces to a discrete Fourier transform [54].

- Wickerhauser [127] proposes an approximation procedure of the Karhunen-Loève transform. A dictionary of approximates is chosen based on a distance given by a transform coding gain metric or by an entropy metric.

- Davila [34] develops a strategy for blind estimation of the Karhunen-Loève basis vectors for wireless communication applications, where the stationarity properties of a signal vary with time.

- Levy and Lindenbaum [65] determine a lower-dimensional Karhunen-Loève basis derived from a sequence of singular value decomposition steps (see [60]). Their strategy applies to the analysis of a sequence of N images of sizes $M \times M$ with N much larger than M.

With respect to bit rate allocation, the Karhunen-Loève transform is still optimal even if the quantizers do not have high bit rates.

2.8 Summary

The Karhunen-Loève transform is one of the most widely used transforms in modern signal processing. It is employed to decorrelate the sub-band signals and also to represent a signal in its principal components. These qualities makes the Karhunen-Loève transform optimal with respect to the coding gain. However, in coding schemes, its properties are optimal only for Gaussian input.

We showed in Sections 2.5 and 2.6 two different properties of the Karhunen–Loève transform. In Section 2.5 we showed how the whole transformed output is used such to insure the encoder works on decorrelated components. The use of the Karhunen–Loève transform based on the property shown in Section 2.6 is almost the opposite: the threshold representation means discarding some of the computed vector basis. This strategy is adopted in computer vision, motion estimation and pattern recognition applications, where the cost to apply the transform – i.e. matrix-vector multiplications – is high, hence lower dimensional solutions are preferable. The dimension reduction is achieved by keeping only the first few vectors of the Karhunen–Loève basis, namely those corresponding to the largest eigenvalues. For details see [103, 82, 78, 81].

The high cost and the data dependence of the Karhunen-Loève trans-

form are significant in the processing and analysis of large collections of images such as hyper-spectral ones. Once we know the performance of this optimal memoryless system, we can search for filter banks that render at least the same coding gain and have good decorrelation properties. The Karhunen-Loève transform is a challenging reference point in the study of filter bank performance and design.

Chapter 3

Optimality of Filter Banks

The purpose of this chapter is to investigate coding gain optimality of orthonormal filter banks. The coding gain is a good indication of the performance of a filter bank in a sub-band coding scheme. The Karhunen-Loéve transform is optimal in the principal component sense – as well as in the coding gain sense. Principal component filter banks – yet to be defined and designed – are desired such that they have the principal component optimality. The rationale for studying in depth the theory and design of coding gain optimal filter banks is first to have a good background on optimality of filter banks for coding schemes and to ensure that principal component filter banks do not violate the coding gain optimality.

3.1   Introduction

A filter bank is shown in Figure 3.1. Generally, the filters are chosen such that the sub-band coders are optimal under certain explicitly stated criteria. Under suitable conditions met by the filter banks considered in this study, the coding gain of a filter bank can be expressed as the ratio of the arithmetic and the geometric means of the sub-band variances that are output

from the analysis filters – see Section 1.5.2. We call the sub-band signals that are output from the analysis bank the "output sub-band signals" because it is the analysis filter bank and its output that give all the information about the performance of the filter bank. The coding gain itself is defined in terms of the output from the analysis filter bank.

The filters are such that they satisfy the *perfect reconstruction* property [117, 106, 125], which is to say that in the absence of quantization errors, the output $\widehat{x}$ is a shifted and scaled version of the input – see Equation (1.11).

3.1.1 Polyphase representation of a filter bank

Recall that the polyphase representation of a single filter was introduced in Section 1.3.2. The polyphase approach can be applied to all the filters in a filter bank. In Figure 3.1 we see that the filtering operation on each channel is applied to the input signal M times. The use of polyphase structures makes the analysis filtering more efficient: the signal is down-sampled first and then processed, hence the channels operate at a lower rate. See Figure 3.2. The use of polyphase structures for filter banks leads to an efficient computational implementation of the basic building blocks shown in Section 1.3.3. Parallel processing is also suitable with the polyphase implementation of a filter bank. From the theoretical point of view, the polyphase approach enables a uniform/unified theoretical treatment of all the filters in the bank at once. It also has the advantage of converting a system that is not time (translation) invariant on scalar-valued signals (Figure 3.1) into a system acting on vector-valued

signals that is time-invariant.

For detailed discussions on polyphase matrices and other fundamental building blocks for multi-rate systems see [117, 85, 84, 75].
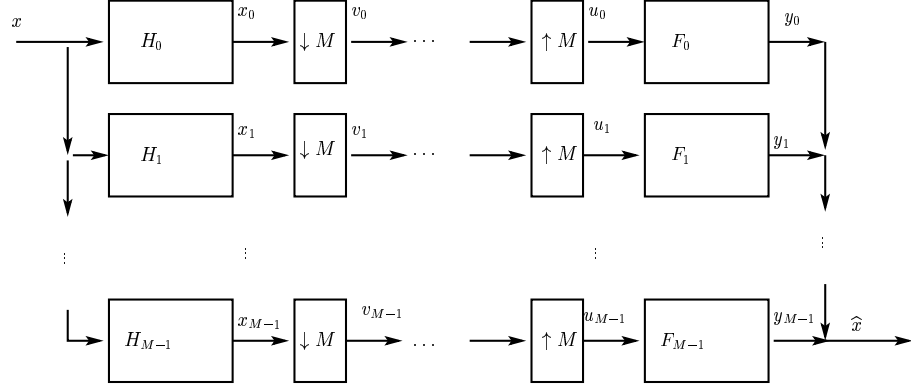
Figure 3.1: Standard form of a uniform maximally down-sampled M-channel filter bank.
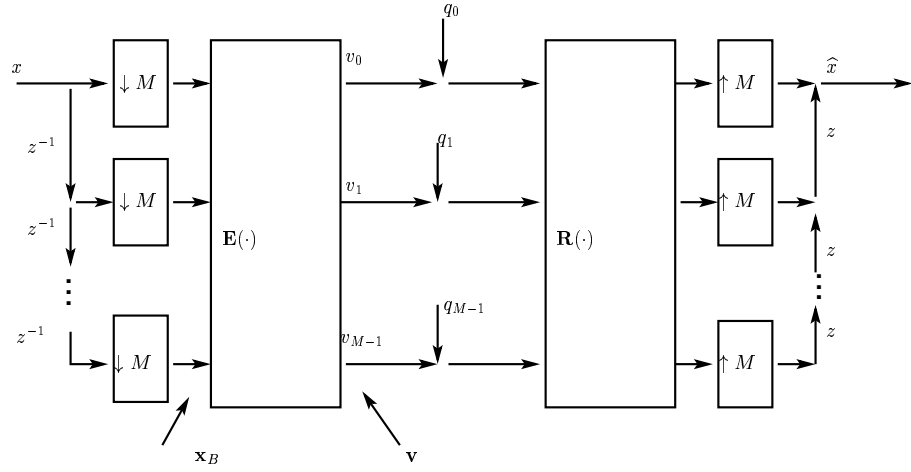
Figure 3.2: The polyphase representation of a filter bank with additive sub-band quantization noise

The analysis polyphase matrix is labeled $\boldsymbol{E}(\cdot)$ and the synthesis polyphase

60

matrix is labeled $\boldsymbol{R}(\cdot)$. The down-samplers and up-samplers have been moved across the polyphase matrices using the noble identities [106]

$$H(z)(\downarrow M)\boldsymbol{x} = (\downarrow M)H(z^M)\boldsymbol{x}\,,$$

$$(\uparrow M)H(z)\boldsymbol{x} = H(z^M)(\uparrow M)\boldsymbol{x}\,.$$

Let the z-transforms of the filters be $H_k(\cdot)$, $k = 0, 1, \ldots, M - 1$. Each of these filters admits a polyphase representation as in Section 1.3.2

$$H_k(z) = \sum_{l=0}^{M-1} z^{-l} E_{kl}(z^M), \qquad k = 0, 1, \ldots, M - 1\,.$$

The polyphase matrix is the matrix $\boldsymbol{E}(z) = [E_{kl}(z)]_{0 \le k, l < M}$. Denote by $\boldsymbol{h}$ the vector whose rows are the frequency response of the analysis filter bank

$$\boldsymbol{h}(z) = \begin{bmatrix} H_0(z) \\ H_1(z) \\ \vdots \\ H_{M-1}(z) \end{bmatrix}, \qquad z \in \mathbb{C}\,.$$

The polyphase matrix and the filter banks are connected via

$$\boldsymbol{h}(z) = \boldsymbol{E}(z^M)\boldsymbol{e}(z), \qquad z \in \mathbb{C}\,,$$

$$\boldsymbol{e}(z) = [1\ z^{-1}\ \ldots\ z^{-(M-1)}]^T\,.$$

The vector $\boldsymbol{e}$ is called a *delay chain.*

Similar to a simple filter expressed in terms of its impulse response, the polyphase matrix of a filter bank can be written as a z-transform of impulse response matrices

$$\boldsymbol{E}(z) = \sum_{k=0}^{K} \boldsymbol{E}_k z^{-k} \tag{3.1}$$

61

Here K is the order of the polyphase matrix (assume that $K \neq 0$). The perfect reconstruction property in terms of polyphase matrices is $\boldsymbol{R}(z)\boldsymbol{E}(z) = \boldsymbol{I}$, and $\widehat{x} = x$. More generally speaking, we can say that a filter bank has perfect reconstruction when $\boldsymbol{E}(\cdot)$ is invertible, because we can always select $\boldsymbol{R}(\cdot)$ to be $\boldsymbol{E}^{-1}(\cdot)$. There are many techniques for the design of perfect reconstruction finite impulse response filter banks [117, 105, 121].

Theorem 3.1.1. [106] *A finite impulse response analysis bank has a finite impulse response synthesis bank that gives perfect reconstruction if and only if the determinant of $\boldsymbol{E}(\cdot)$ is a monomial:*

$$\det \boldsymbol{E}(z) = cz^{-l}, \qquad l \in \mathbb{Z},$$

for some nonzero constant c.

The polyphase representation of a filter bank enables us to treat the filtering of a signal as a multi-input-multi-output linear time-invariant system.

3.1.2   Uniform filter banks

Filter banks can be divided into two classes based on the way the signal is blocked and sent though the channels: *uniform* when all the channels are down-sampled by the same factor and *nonuniform* otherwise. A uniform *maximally down-sampled* M-channel filter bank has each channel down-sampled by M, whereas a maximally down-sampled nonuniform M-channel filter bank satisfies $\sum_{k=0}^{M-1} \frac{1}{n_k} = 1$ where each natural number n_k is the down-sampled ratio on each channel. Of the two classes, uniform filter banks are the more

common in practice, as they are easier both to analyze and to implement than the nonuniform ones.

Uniform filter banks can be classified in terms of their analysis and synthesis polyphase matrices $\boldsymbol{E}(\cdot)$ and $\boldsymbol{R}(\cdot)$, respectively (See Figure 3.2). The types of uniform filter banks and their characteristics are listed below.

- Biorthogonal:

  $\boldsymbol{R}(z)\boldsymbol{E}(z) = \boldsymbol{I}$. The filter transfer functions satisfy:

  $H_k(z)F_j(z)|_{\downarrow M} = \delta_{kj}$ where

  $X(z)|_{\downarrow M}$ is the z-transform of the down-sampled sequence $\{x[nM]\}_{n\in\mathbb{Z}}$.

- Orthonormal or para-unitary:

  $\boldsymbol{R}(z) = \widetilde{\boldsymbol{E}}(z) = \boldsymbol{E}_*^T(z^{-1})$, the matrices $\boldsymbol{E}(\cdot)$ and $\boldsymbol{R}(\cdot)$ are para-unitary

  $F_k(z) = \widetilde{H}_k(z)$ the z-transform of $h^*[-n]$

  $f_k[n] = h_k^*[-n]$

- Transform coder: a filter bank such that the polyphase matrix is constant: $\boldsymbol{E}(z) = \boldsymbol{E}_0,\ z \in \mathbb{C}$. The filter lengths do not exceed the number of channels.

- Finite impulse response filter bank: a filter bank with finite impulse response filters; in this case the polyphase matrix can be written as: $\boldsymbol{E} = \boldsymbol{E}_0 + \boldsymbol{E}_1 z^{-1} + \cdots + \boldsymbol{E}_K z^{-K}$, where K is the order of the multiple input multiple output system

63

- Infinite impulse response filters – see Section 1.3. Among these there are the ideal filter banks – see Figure 1.4.

3.1.3 The McMillan degree of a filter bank

Recall the order of a filter defined in Section 1.3. The order of a filter shows the number of delays required to implement the filter and it also shows how much input memory is used to produce each output sample given – see Equation (1.7)

$$y[n] = \sum_{k=0}^{N} h[k]x[n-k] \,.$$

The order of a filter is a measure of its complexity.

A filter bank can be regarded as a multi-input-multi-output system and the transfer matrix of the system is the polyphase matrix of the filter bank. The polyphase matrix $\boldsymbol{E}$ can be written as in Equation (3.1)

$$\boldsymbol{E}(z) = \sum_{k=0}^{K} \boldsymbol{E}_k z^{-k} \,.$$

The complexity of a filter bank, as measured by the number of delays required to implement it, is not given by K. To illustrate this point, consider that the simple first-order polyphase matrix $\boldsymbol{E}(z) = z^{-1}\boldsymbol{I}$ of an M-channel filter bank requires M delays to implement it – see Figure 3.2. This is the minimum number of delays required to implement the system and this number, M, is called the McMillan degree of the system. Non-causal systems cannot be implemented by delays alone, hence the definition of McMillan degree is restricted to causal systems.

In order to show how the McMillan degree is defined in general, a few preliminary results are necessary.

Let M and N be two positive integers. A matrix polynomial $\boldsymbol{H}(z) \in \mathbb{C}^{M \times N}[z]$ is a $M \times N$ matrix whose entries are polynomials in the complex variable z

$$\boldsymbol{H}(z) = \sum_{n=0}^{K} \boldsymbol{h}_n z^n, \qquad K \geq 0 .$$

The order of a polynomial matrix is the largest number K that satisfies $\boldsymbol{h}_K \neq 0$. The notion of rank can be associated with a matrix polynomial and the rank depends on the value of the variable z; in general $rank(\boldsymbol{H}(z)) \leq \min(M, N)$. Specific to matrix polynomials is the *normal rank* defined next [42, 117].

Definition 3.1.1. Let $\boldsymbol{H} \in \mathbb{C}^{M \times N}[z]$ be a polynomial matrix. The normal rank of $\boldsymbol{H}(z)$ is

$$\rho(\boldsymbol{H}(z)) = \max_{z \in \mathbb{C}} rank(\boldsymbol{H}(z)) . \tag{3.2}$$

A matrix polynomial has *full rank* when its normal rank is maximum, i.e. $\rho(\boldsymbol{H}(z)) = \min\{M, N\}$.

Definition 3.1.2. A square matrix polynomial $\boldsymbol{U} \in \mathbb{C}^{M \times M}[z]$ with constant non-zero determinant is called *unimodular*.

The normal rank of an $M \times M$ unimodular matrix is M and the inverse of a unimodular matrix is also unimodular.

The next result gives the representation of a polynomial matrix. Let $\boldsymbol{H}(z)$ be a polynomial matrix of size $M \times N$.

65

Theorem 3.1.2. [42, 53] *(Smith Form of a Polynomial Matrix). Every matrix polynomial $\boldsymbol{H}(z)$ with normal rank q admits the representation*

$$\boldsymbol{H}(z) = \boldsymbol{P}(z)\boldsymbol{D}(z)\boldsymbol{Q}(z) \tag{3.3}$$

where

$$\boldsymbol{D}(z) \;=\; \begin{bmatrix} d_1(z) & & \cdots & & 0 \\ & & \ddots & & \\ \vdots & & d_q(z) & & \vdots \\ & & & 0 & \\ & & & & \ddots \\ 0 & & \cdots & & 0 \end{bmatrix} \tag{3.4}$$

is a diagonal matrix polynomial with monic scalar polynomials $d_k(z)$ such that $d_k(z)$ is divisible by $d_{k-1}(z)$. The matrices $\boldsymbol{P}(z)$ and $\boldsymbol{Q}(z)$ are unimodular matrices.

The matrix $\boldsymbol{D}(z)$ in the Smith representation is unique, although $\boldsymbol{P}$ and $\boldsymbol{Q}$ are not.

Let $\boldsymbol{E}(z)$ be the polyphase matrix of a causal finite impulse response filter bank of order K Rather than using the Smith form of $\boldsymbol{E}(z)$ with respect to the variable z^{-1} (i.e. regarding $\boldsymbol{E}(z)$ as a polynomial matrix in z^{-1}), the McMillan degree of $\boldsymbol{E}(z)$ is instead based on a closely related canonical from known as the Smith-McMillan form, defined as follows:

1. Factor z^{-K} out of $\boldsymbol{E}(z)$ and write this in the form

$$\boldsymbol{E}(z) = z^{-K}\boldsymbol{F}(z)$$

where $\boldsymbol{F}(z)$ is a matrix polynomial in z.

2. Let

$$\boldsymbol{F}(z) = \boldsymbol{P}(z)\boldsymbol{D}(z)\boldsymbol{Q}(z)$$

be the Smith representation of $\boldsymbol{F}(z)$.

3. The Smith-McMillan representation of $\boldsymbol{E}(z)$ is then defined to be

$$\boldsymbol{F}(z) = \boldsymbol{P}(z)\boldsymbol{\Lambda}(z)\boldsymbol{Q}(z)$$

where

$$\boldsymbol{\Lambda}(z) = z^{-K}\boldsymbol{D}(z)$$

is the diagonal Smith-McMillan form for $\boldsymbol{F}(z)$.

The Smith-McMillan invariant factors for $\boldsymbol{E}(z)$ (the diagonal entries in $\boldsymbol{\Lambda}(\mathrm{z})$) have the form

$$l_k(z) = \frac{d_k(z)}{z^K}$$

where the polynomials $d_k(z)$ are the Smith invariant factors for $\boldsymbol{F}(z)$.

Since $\boldsymbol{P}(z)$ and $\boldsymbol{Q}(z)$ are unimodular, the perfect reconstruction condition

$$\det \boldsymbol{E}(z) = cz^{-d}, \qquad d \geq 0, \; c \neq 0 \tag{3.5}$$

implies that the numerators (Smith invariant factors) are monomials:

$$d_k(z) = z^{m_k}, \qquad m_k \geq 0\,.$$

This means that the Smith-McMillan invariant factors for $\boldsymbol{E}(z)$ look like

$$l_k(z) = z^{m_k - K}\,.$$

67

It is possible to be more precise about the size of the exponents for the Smith-McMillan invariant factors in the para-unitary case. In particular, when $\boldsymbol{E}(z)$ is para-unitary, Vaidyanathan [117] and Brislawn [19] show that

$$0 \leq m_k \leq K$$

Thus the Smith-McMillan form for $\boldsymbol{E}(z)$ is

$$\boldsymbol{\Lambda}(z) = \begin{bmatrix} z^{-n_0} & 0 & \cdots & 0 \\ 0 & z^{-n_1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & z^{-n_{M-1}} \end{bmatrix}$$

with $n_0 \geq n_1 \geq \cdots \geq n_{M-1} \geq 0$. Note that $\boldsymbol{\Lambda}(z)$ is a causal system when $\boldsymbol{E}(z)$ is para-unitary. When $\boldsymbol{E}(z)$ is causal para-unitary finite impulse response, the McMillan degree, μ, is defined to be

$$\mu = \sum_{k=0}^{M-1} n_k \,.$$

Also, taking the determinant of the Smith-McMillan representation from Equation (3.5) shows that

$$\deg(\det \boldsymbol{E}(z)) = \mu$$

In general for any causal system, the McMillan degree is defined by a generalization of the above procedure. It is always a nonnegative integer, although it is not always the same as the polynomial degree of $\det \boldsymbol{E}(z)$. By a theorem of Kalman [57], there exists an implementation of $\boldsymbol{E}(z)$ with exactly μ delays and there are no implementations with fewer delays. For causal finite impulse response systems, the McMillan degree is less than or equal to

the order of the system. For example, transform coders are memoryless (order zero) filter banks with McMillan degree zero. The first order polyphase matrix $\boldsymbol{E}(z) = z^{-1}\boldsymbol{I}$ of an M-channel filter bank has McMillan degree M greater than 1, however.

The representation of polyphase matrices in Smith-McMillan form is useful in both analysis and design of filter banks and their properties.

3.2 Modeling Considerations

3.2.1 Orthonormal filter banks

The present theory is applicable to the "universe" of all orthonormal filter banks. In practice, however, finite impulse response filter banks are preferred over infinite impulse response ones as they are typically easier to analyze and to implement. Also, finite impulse response filters lead to compactly supported wavelets and so enjoy the benefits of multi-resolution analysis [31, 4, 114, 124, 125].

The number of channels M is fixed greater than one. The terms "subband" and "channel" are used interchangeably.

Whenever possible, the analysis filter banks are assumed causal. Note that ideal filter banks are non-causal. Orthonormal or para-unitary filter banks have the analysis polyphase matrix $\boldsymbol{E}(\cdot)$ unitary at all frequencies,

$$\widetilde{\boldsymbol{E}}(e^{i\omega})\boldsymbol{E}(e^{i\omega}) = \boldsymbol{I}, \qquad \omega \in [0, 2\pi) . \tag{3.6}$$

If $\boldsymbol{E}(z)$ is causal and para-unitary then the synthesis filter bank is anti-causal

(i.e. polynomial in z) by the definition of $\widetilde{\boldsymbol{E}}(z)$. The polyphase matrix of a causal orthonormal filter bank with McMillan degree μ can be represented as a cascade of generalized Householder based matrices and a unitary matrix $\boldsymbol{U}$ – see Vaidyanathan [117]:

$$\boldsymbol{E}(z) = \boldsymbol{V}_0(z)\boldsymbol{V}_1(z)\dots\boldsymbol{V}_{\mu-1}(z)\boldsymbol{U}, \tag{3.7}$$

$$\boldsymbol{V}_k(z) = \boldsymbol{I} + (-1 + z^{-1})\boldsymbol{v}_k\boldsymbol{v}_k^T, \qquad \|\boldsymbol{v}_k\| = 1,\ 0 \le k < \mu. \tag{3.8}$$

This representation is based on dyadic structures $\boldsymbol{I} + (-1 + z^{-1})\boldsymbol{v}\boldsymbol{v}^T$ that are Householder matrices when z is -1; the vector $\boldsymbol{v}$ has unit norm.

As a generalization of the transform coder, a filter bank is considered optimal if it satisfies the same optimality criterion, i.e. it maximizes the coding gain or equivalently minimizes the mean-square reconstruction error due to sub-band quantization.

In an orthonormal filter bank, each filter satisfies the Nyquist(M) constraint

$$|H_k(e^{i\omega})|^2_{\downarrow M} = 1, \qquad 0 \le k < M \tag{3.9}$$

or equivalently

$$\sum_{l=0}^{M-1} |H_k(e^{i(\omega-2\pi l/M)})|^2 = M, \qquad 0 \le k < M \tag{3.10}$$

The optimal filters we consider are alias-free(M) or anti-alias(M) filters, i.e. the output of the filter can be down-sampled without aliasing or equivalently the shifted versions $H_l(e^{i(\omega-2k\pi/M)})$ do not overlap for $k \ne l$, $k, l \in$

70
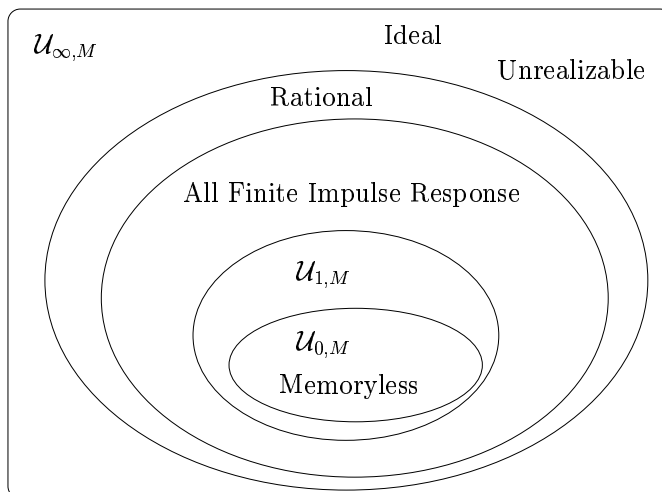
Figure 3.3: The set of orthonormal filter banks

$\{0, 1, ..., M-1\}$ – see Section 1.3.3. We say that these filters have alias-free(M) support and for such filters there could be multiple pass-bands.

The "universe" of orthonormal filter banks is shown in Figure 3.3. Denote by $\mathcal{U}_{\infty,M}$ the set of maximally down-sampled uniform M-channel orthonormal filter banks with unrestricted orders. The "universe" $\mathcal{U}_{\infty,M}$ contains all of the finite impulse response orthonormal filter banks; the sets of finite impulse response filter banks with McMillan degree μ and the class of transform coders are denoted by $\mathcal{U}_{\mu,M}$ and $\mathcal{U}_{0,M}$, respectively. Among the infinite impulse response filter banks, $\mathcal{U}_{\infty,M}$ contains the ideal passband filters and infinite impulse response rational filter banks.

3.2.2   Statistical models

This section gives the models for the input to orthonormal filter banks and shows the action of the filters on the input signal. Also, the model for the quantization error and the definition of the optimality of filter banks in the presence of quantizer errors are given.

The input signal x is a zero mean, wide-sense stationary stochastic process with given power spectral density $S_{xx}(\cdot)$ and variance σ_x^2. The input signal passes through a filter bank that outputs the signals $v_k, \ k = 0, 1, \ldots, M-1$ called the *output sub-band signals*. The sub-band signals $v_k(\cdot)$ have zero mean and are jointly wide sense cyclo-stationary with variances $\sigma_{v_k}^2$. The term "output" refers to the output from the analysis bank. In the polyphase representation the input signal is blocked and the output signals v_k are the components of the output vector signal $\boldsymbol{v} = [v_0, v_1, \ldots, v_{M-1}]^T$ – see Figure 3.2

The power spectral density matrix of the output signal is given by

$$\boldsymbol{S_{vv}}(e^{i\omega}) = \boldsymbol{E}(e^{i\omega})\boldsymbol{S_{xx}}(e^{i\omega})\boldsymbol{E}^\dagger(e^{i\omega}), \qquad \omega \in [0, 2\pi] \qquad (3.11)$$

where $\boldsymbol{E}(\cdot)$ is the polyphase matrix of the filter bank and the output sub-band variance vector, $\boldsymbol{\sigma_v^2}$, and its components are

$$\begin{aligned}
\boldsymbol{\sigma_v^2} &= [\sigma_{v_0}^2, \ \sigma_{v_1}^2, \ldots, \sigma_{v_{M-1}}^2]^T \text{and} \\
\sigma_{v_k}^2 &= \int_0^{2\pi} [\boldsymbol{S_{vv}}(e^{i\omega})]_{kk} d\omega/(2\pi) .
\end{aligned} \qquad (3.12)$$

By convention, the filters output the sub-band variances ordered by size, so that

$$\sigma_{v_0}^2 \geq \sigma_{v_1}^2 \geq \ldots \geq \sigma_{v_{M-1}}^2 .$$

72

In any filter bank this ordering can be achieved via a permutation of the filters without affecting the perfect reconstruction or the orthonormality properties.

Proposition 3.2.1. *1) A uniformly down-sampled wide-sense stationary input is also wide-sense stationary and the variance of the down-sampled signal equals the input variance.*
2) The average of the output sub-band variances from a uniform maximally down-sampled orthonormal filter bank is equal to the input variance.

Proof. 1) Suppose the input signal is down-sampled uniformly by a factor of M; the down-sampled sequence $(\downarrow M)x$ is described by

$$(\downarrow M)x[n] = x[nM], \qquad n \in \mathbb{Z}.$$

The mean of the down-sampled sequence is

$$E\{(\downarrow M)x[n]\} = E\{x[nM]\} = \mu_x = 0, \qquad n \in \mathbb{Z},$$

and the autocorrelation is

$$R_M[n, n-l] = E\{(\downarrow M)x[n]\} \, ((\downarrow M)x[n-l])^*\} = R[lM], \qquad n, l \in \mathbb{Z}.$$

I only depends on the lag l, and, therefore, the down-sampled signal is also wide-sense stationary. The variance of the down-sampled signal equals the input variance since $R_M[n, n] = R[0]$.

2) Denote the trace of the output power spectral density matrix by t, where

$$t = \text{tr } \boldsymbol{S_{vv}}(e^{i\omega}) = \sum_{k=0}^{M-1} [\boldsymbol{S_{vv}}(e^{i\omega})]_{kk}.$$

73

Since $\boldsymbol{E}(e^{i\omega})$ is unitary at each ω, we deduce from Equation (3.11) that t is constant and equals the trace of the input power spectral density, since for any matrix $\boldsymbol{A} \in \mathbb{R}^M \times M$ and any unitary matrix $\boldsymbol{U} \in \mathbb{C}^M \times M$ th trace satisfies

$$\operatorname{tr} \boldsymbol{A} = \operatorname{tr} \boldsymbol{EAE}^{\dagger}.$$

Equation (3.12) shows that t is invariant under the integration, and so the sum of the output variances is constant and averages the input variance. $\square$

Proposition 3.2.2. *Let a signal be such that its power spectral density does not vanish on a set of positive measure in $[0, 2\pi]$ and let this signal be the input to a para-unitary filter bank. Then the output sub-band variances are strictly positive. Furthermore, the coding gain is well defined.*

Proof. [88, 101]. Let the number of channels be M. Refer to Figure 3.1; since for each filter in the bank,

$$S_{v_k v_k}(e^{i\omega}) = S_{xx}(e^{i\omega})|H_k(e^{i\omega})|^2, \qquad k = 0, 1, \ldots, M-1,$$

and the filter bank is orthonormal, it follows that the sub-band variances are nonzero, otherwise the orthonormality of the filter bank would be contradicted. The output sub-band variances are nonnegative by the Wiener-Khinchin criterion [38, 37, 33]. $\square$

Quantizer Noise. The model for the quantization error is the standard noise model from Section 1.5.2. The quantizer is modeled by additive

noise sources $q_k(\cdot)$ that are random processes, jointly wide sense stationary with zero-mean and variances

$$\sigma_{q_k}^2 = c2^{-2b_k}\sigma_{x_k}^2 = c2^{-2b_k}\sigma_{v_k}^2, \qquad k = 0, 1, \ldots, M-1$$

Here b_k is the bit budget allowed for the sub-band k and $b = \sum_{k=0}^{M-1} b_k/M$ is the fixed average bit rate. The constant c is associated only with the probability density function of the quantizer's input. The standard noise model conveys that signals with higher variance have more bits allocated when quantization is involved. The noise sources $q_k(\cdot)$ are not constrained to be white or have any two noise sources uncorrelated.

In many signal processing tasks the performance of a filter bank is reflected by two *gains* expressed in terms of the output sub-band variances. These are the *coding gain* and the *compaction gain*. Recall the formula for the coding gain from Section 1.5.2

$$G_{coding} = \frac{(1/M)\sum_{k=0}^{M-1} \sigma_{v_k}^2}{(\prod_{k=0}^{M-1} \sigma_{v_k}^2)^{1/M}} = \frac{\sigma_x^2}{(\prod_{k=0}^{M-1} \sigma_{v_k}^2)^{1/M}}. \tag{3.13}$$

The compaction gain is defined as

$$G_{comp} = \frac{\sigma_{v_0}^2}{\sigma_x^2} \tag{3.14}$$

with the convention that $\sigma_{v_0}^2 \geq \sigma_{v_1}^2 \geq ... \geq \sigma_{v_{M-1}}^2$.

A definition of optimality. With these assumptions on the filter banks and input data and noise models as stated above, a filter bank is optimal for coding gain over a certain class of filters $\mathcal{C} \subset \mathcal{U}_{\infty,M}$ if its coding

75

gain is maximized over $\mathcal{C}$. Equivalently, the product of sub-band variances is minimized over $\mathcal{C}$.

Unless otherwise specified, optimality of a filter bank is understood in the coding gain sense. The coding gain is a positive valued function defined on $\mathcal{U}_{\infty,M}$ (Figure 3.3). The lowest value of the coding gain is 1, from the arithmetic/geometric mean inequality.

Definition 3.2.1. Let x be the input to an orthonormal uniform filter bank with an arbitrary finite number of channels, M. The down-sampled signals v_k, $k = 0, 1, ..., M - 1$ are *totally decorrelated* if, for all integers m, n

$$E\{v_k[n]v_l^*[m]\} = c_l^2 \delta_{kl}\delta_{mn}, \qquad k, l = 0, 1, \ldots, M - 1\,, \qquad (3.15)$$

The down-sampled sub-band signals are decorrelated as random processes, not only as random variables; in the former case we have, for each fixed integer n

$$E\{v_k[n]v_l^*[n]\} = c_l^2 \delta_{kl} \qquad \text{for } k \neq l\,. \qquad (3.16)$$

Note that in the previous relations the value of the correlation c_l^2 depends only on the component index, not on the "time" value; this simplification is justified by linearity of the filters and the property of the input of being a wide sense stationary process.

Denote by $S_k(\cdot)$ the power spectral density of the k-th down-sampled sub-band signal $v_k(\cdot)$ and let $\sigma_{v_k}^2$ be its variance.

Definition 3.2.2. Assume that the output sub-bands variances are ordered decreasingly

$$\sigma_{v_0}^2 \geq \sigma_{v_1}^2 \geq ... \geq \sigma_{v_{M-1}}^2 .$$

The sub-band signals, or the sub-band power spectra $\{S_k(\cdot)\}_{0 \leq k \leq M-1}$ have the *spectral majorization* property if

$$S_0(e^{i\omega}) \geq S_1(e^{i\omega}) \geq \ldots \geq S_{M-1}(e^{i\omega}), \qquad \omega \in [0, 2\pi]. \qquad (3.17)$$

This property gives information about the action of the filters upon the input and is a key ingredient in the study of optimality with respect to the coding gain.

3.3   Coding Gain

The results here show when a filter bank is optimal with respect to the coding gain in $\mathcal{U}_{\infty,M}$. See [114, 118].

In Chapter 2 it is shown that the orthogonal transform coder with a constant polyphase matrix $\boldsymbol{E} \in \mathbb{C}^{M \times M}$ maximizes the coding gain over all the transform coders if and only if it decorrelates the input signal – see Equation (3.16). In the case of a sub-band coder stronger conditions are necessary for coding gain optimality over $\mathcal{U}_{\infty,M}$.

3.3.1   Necessary conditions for optimality

In the next results the coding gain optimality is considered over the whole universe $\mathcal{U}_{\infty,M}$.

Theorem 3.3.1. *The sub-band signals of a coding gain optimal filter bank over $\mathcal{U}_{\infty,M}$ satisfy the total decorrelation property as in Definition 3.2.1.*

Proof. [118] Consider an optimal filter bank over $\mathcal{U}_{\infty,M}$. Suppose a pair of sub-band signals are correlated. It suffices to show that the coding gain can be increased without violating the orthonormality of the filter bank. Without loss of generality we may assume that the first two sub-band signals are such that the random processes $\boldsymbol{v}_0$ and a shifted version of $\boldsymbol{v}_1$ are correlated, that is, there exists an integer k such that for all integers n,

$$E\{v_0[n]v_1^*[n-k]\} \neq 0\,.$$

Since the input is wide sense stationary, the inequality holds for all integers n. Let $\boldsymbol{\Theta}$ be the Karhunen-Loève transform matrix for the vector process $[v_0[n] \ v_1^*[n-k]]^T$. An analysis filter bank can be constructed to decorrelate v_0 and v_1, using $\boldsymbol{\Theta}$ and a delay z^{-k}. The polyphase matrix of this new filter bank is

$$\begin{aligned}
\boldsymbol{E}_{\Theta,k}(z) &= \begin{bmatrix} \boldsymbol{\Theta} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I} \end{bmatrix} \boldsymbol{\Lambda}(z)\boldsymbol{E}(z), \quad \text{with} \\
\boldsymbol{\Lambda} &= \operatorname{diag}\{1,\ z^{-k},\ 1,\ \ldots,\ 1\}\,.
\end{aligned}$$

The matrices $\boldsymbol{E}, \boldsymbol{\Lambda}$ and $\boldsymbol{\Theta}$ are para-unitary, hence their product is a para-unitary matrix $\boldsymbol{E}_{\Theta,k}$. Let $w_k,\ k = 0,1,\ldots,M-1$ be the sub-band signals from the new filter bank with polyphase matrix $\boldsymbol{E}_{\Theta,k}$, as in Figure 3.4 and let $\boldsymbol{R_w}$ and $\boldsymbol{R_v}$ be the correlation matrices of the vectors $[w_0[n] \ w_1^*[n]]^T$ and

78
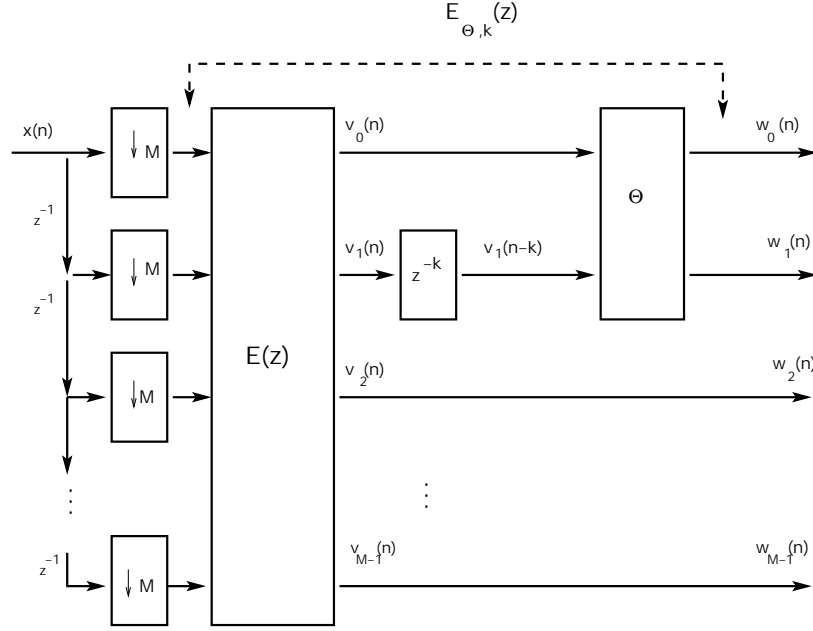
Figure 3.4: Coding gain increased by a decorrelation transform

$[v_0[n]\ v_1^*[n-k]]^T$ respectively. From the definition of $\boldsymbol{\Theta}$ it follows that

$$\boldsymbol{R_w} = \boldsymbol{\Theta R_v \Theta}^\dagger$$

and since $\boldsymbol{\Theta}$ is a Karhunen-Loève transform, the matrix $\boldsymbol{R_w}$ is diagonal. Since $\boldsymbol{\Theta}$ is unitary it follows that

$$\det(\boldsymbol{R_w}) = \det(\boldsymbol{R_v})\,.$$

Hadamard inequality yields:

$$\sigma_{w_0}^2 \sigma_{w_1}^2 < \sigma_{v_0}^2 \sigma_{v_1}^2$$

79

where $\sigma^2_{w_k}, \sigma^2_{v_k}$, $k = 0, 1$ are the diagonal elements of the matrices $\boldsymbol{R_w}$ and $\boldsymbol{R_v}$, respectively [60, 15]. The inequality $\sigma^2_{w_0}\sigma^2_{w_1} < \sigma^2_{v_0}\sigma^2_{v_1}$ means that the coding gain is greater for the polyphase matrix $\boldsymbol{E_{\Theta,k}}$ which contradicts the fact that the filter bank with polyphase $\boldsymbol{E}$ is optimal in $\mathcal{U}_{\infty,M}$. In conclusion, the total decorrelation of the sub-band signals – as stochastic processes – is a necessary condition for coding gain optimality. $\qquad\qquad\square$

The above theorem has another interpretation. When the analysis filter bank is optimal with respect to the coding gain, the power spectral density matrix of the vector stochastic process $\boldsymbol{v} = [v_0 \ v_1 \ \cdots \ v_{M-1}]^T$ of sub-band signals is diagonal

$$\boldsymbol{S_{vv}}(e^{i\omega}) = \begin{bmatrix} S_0(e^{i\omega} & 0 & \cdots & 0 \\ 0 & S_1(e^{i\omega}) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & S_{M-1}(e^{i\omega}) \end{bmatrix}, \qquad \omega \in [0, 2\pi)$$

with $S_k(\cdot)$ being the power spectral density of sub-band signal v_k with $k = 0, 1, \ldots, M - 1$. This is because correlation of sub-bands translates into non-zero terms off the diagonal in the power spectral density matrix.

The next theorem shows another necessary condition for optimality of a filter bank. Again the optimality is over $\mathcal{U}_{\infty,M}$ and it is understood in the coding gain sense: a filter bank is optimal if its coding gain is maximum in $\mathcal{U}_{\infty,M}$.

Theorem 3.3.2. *The output sub-band signals of a coding gain optimal filter bank over $\mathcal{U}_{\infty,M}$ satisfy the spectral majorization property as in Definition 3.2.2.*

Proof. [118] Suppose that the majorization property is not satisfied. Assume without loss of generality that the first two sub-band spectra are not majorized, $S_0(\cdot)$ and $S_1(\cdot)$. Then, there are values of the argument ω at which $S_0(e^{i\omega}) \geq S_1(e^{i\omega})$ does not hold; an example is shown in Figure 3.5.
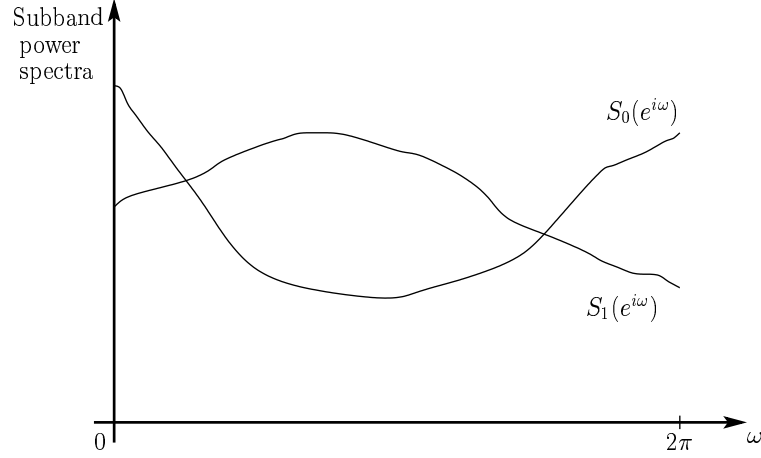
Figure 3.5: Output spectra without the spectral majorization property.

Let $\boldsymbol{E}$ be the polyphase matrix of the original analysis filter bank. Construct a matrix $\boldsymbol{T}$ as

$$\boldsymbol{T}(e^{i\omega}) = \begin{cases} \boldsymbol{I}_M, & \text{if } S_0(e^{i\omega}) \geq S_1(e^{i\omega}) \\ \begin{bmatrix} 0 & 1 & \boldsymbol{0} \\ 1 & 0 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I}_{M-2} \end{bmatrix}, & \text{if } S_0(e^{i\omega}) < S_1(e^{i\omega}) \end{cases}$$

where $\boldsymbol{I}_M$ and $\boldsymbol{I}_{M-2}$ are the identity matrices in $\mathbb{R}^M$ and $\mathbb{R}^{M-2}$, respectively. Cascade the matrix $\boldsymbol{T}$ as in Figure 3.6. Denote by $S_{T,0}(\cdot)$ and $S_{T,1}(\cdot)$ the pair of output spectra of the first sub-band, after applying $\boldsymbol{T}$. For all ω in $[0, 2\pi]$

81
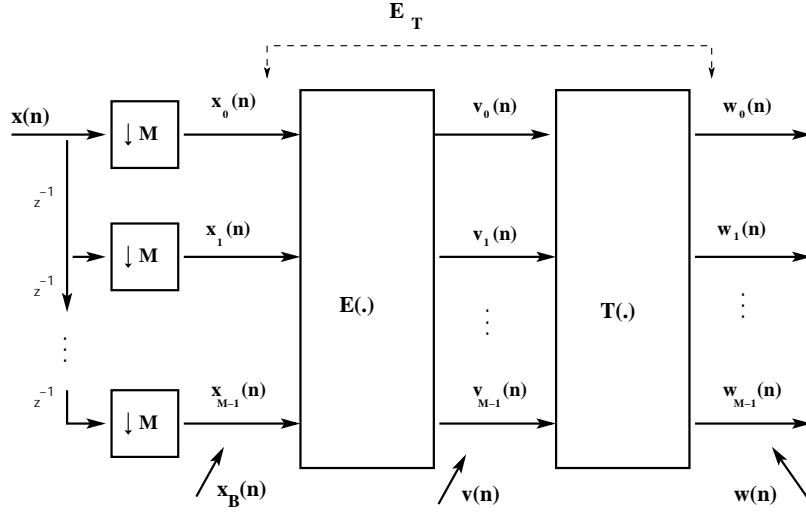
Figure 3.6: The new polyphase matrix for sub-band decorrelation.

the following inequalities hold;

$$
\begin{aligned}
S_{T,0}(e^{i\omega}) &\geq S_{T,1}(e^{i\omega}), \\
S_{T,0}(e^{i\omega}) &\geq S_0(e^{i\omega}), \\
S_{T,1}(e^{i\omega}) &\leq S_1(e^{i\omega}).
\end{aligned}
$$

Thus the variances of the new sub-band signals w_0 and w_1 are such that

$$
\sigma_{w_0}^2 \geq \sigma_{v_0}^2 \text{ and } \sigma_{w_1}^2 \leq \sigma_{v_1}^2 \ .
$$

The analysis polyphase matrix $\boldsymbol{E}_T = \boldsymbol{T}\boldsymbol{E}$ is para-unitary since $\boldsymbol{T}$ is unitary

by construction; hence the sum of output variances is unchanged,

$$\sigma_{w_0}^2 + \sigma_{w_1}^2 = \sigma_{v_0}^2 + \sigma_{v_1}^2 \ .$$

Thus the new variances can be written in terms of the old ones

$$\sigma_{w_0}^2 = \sigma_{v_0}^2 + \delta, \ \ \sigma_{w_1}^2 = \sigma_{v_1}^2 - \delta, \qquad \delta > 0 \ .$$

Evaluate the product of the new sub-band variances

$$\sigma_{w_0}^2 \sigma_{w_1}^2 = \sigma_{v_0}^2 \sigma_{v_1}^2 - \delta^2 - \delta(\sigma_{v_0}^2 - \sigma_{v_1}^2) < \sigma_{v_0}^2 \sigma_{v_1}^2$$

The coding gain is thus increased and that contradicts the optimality over $\mathcal{U}_{\infty,M}$ of the filter bank with polyphase matrix $\boldsymbol{E}$. $\qquad\square$

3.3.2   Sufficient conditions for optimality

Neither of the necessary conditions from the previous two theorems is sufficient by itself for coding gain optimality, but if they are both satisfied, then optimality over $\mathcal{U}_{\infty,M}$ is attained.

Theorem 3.3.3. *(Necessary and sufficient conditions for optimality.) Let a signal with given power spectral density be input to filter banks in $\mathcal{U}_{\infty,M}$. The coding gain is maximized over $\mathcal{U}_{\infty,M}$ if and only if the down-sampled output sub-band signals satisfy the total decorrelation and the spectral majorization properties. Moreover, the set of down-sampled sub-band power spectra is unique – but the optimal orthonormal filter bank may not be unique.*

Proof. [118] This proof relies on the existence of optimal filter banks over $\mathcal{U}_{\infty,M}$. Let $\boldsymbol{E}(\cdot)$ provide total decorrelation of sub-band spectra and spectral majorization. From the total decorrelation property of $\boldsymbol{E}(\cdot)$ it follows that the output power spectral density matrix is diagonal at each frequency $\omega \in [0, 2\pi]$,

$$
\begin{aligned}
\boldsymbol{S_{vv}}(e^{i\omega}) &= \boldsymbol{E}(e^{i\omega})\boldsymbol{S_{xx}}(e^{i\omega})\boldsymbol{E}^\dagger(e^{i\omega}) \\
&= \text{diag}\{S_0(e^{i\omega}) \ S_1(e^{i\omega}) \ \dots \ S_{M-1}(e^{i\omega})\}.
\end{aligned}
\tag{3.18}
$$

The matrix $\boldsymbol{E}(\cdot)$ is unitary at each frequency, hence the output sub-band spectra $S_0(e^{i\omega})$, $S_1(e^{i\omega}), \dots, S_{M-1}(e^{i\omega})$ are the eigenvalues of the input power spectral density matrix calculated at each $\omega \in [0, 2\pi]$. At each frequency, the set of eigenvalues is unique. Let $\boldsymbol{F}(\cdot)$ be a polyphase matrix that is known to maximize the coding gain. The polyphase matrix $\boldsymbol{F}(\cdot)$ diagonalizes the input power spectral density matrix, at each frequency, whence

$$
\begin{aligned}
\boldsymbol{S}_{\boldsymbol{vv}}^{(F)}(e^{i\omega}) &= \boldsymbol{F}(e^{i\omega})\boldsymbol{S_{xx}}(e^{i\omega})\boldsymbol{F}^\dagger(e^{i\omega}) \\
&= \text{diag}\{S_0^{(F)}(e^{i\omega}) \ S_1^{(F)}(e^{i\omega})\dots S_{M-1}^{(F)}(e^{i\omega})\}.
\end{aligned}
$$

The set of eigenvalues of a matrix is unique up to a permutation. The spectral majorization property of $\boldsymbol{E}(\cdot)$ guarantees that at each frequency

$$
S_k(e^{i\omega}) = S_k^{(F)}(e^{i\omega}), \qquad k = 0, 1, \dots, M-1.
$$

Hence the coding gain of $\boldsymbol{E}(\cdot)$ is also maximum over $\mathcal{U}_{\infty,M}$. However, the set of eigenvectors of a matrix are not necessarily unique, therefore the analysis polyphase matrix $\boldsymbol{E}$, and hence the filter bank may not be unique for a given input power spectral density. $\qquad\square$

3.4 Energy Compaction

3.4.1 Compaction filters

Definition 3.4.1. An analysis filter $H(\cdot)$ whose output variance σ_v^2 is maximized subject to the constraint that $|H(\cdot)|^2$ be Nyquist(M) [118] is called an *optimum compaction filter.*

An M-channel orthonormal filter bank is an optimum compaction filter bank if it contains one optimum compaction filter. The compaction gain defined in Equation (3.14) shows the ability of a filter bank to pack most of its output energy in one channel. This is a useful property in applications such as signal analysis, denoising, compression and progressive transmission [6, 63]. The compaction gain is an indicator of the performance of a filter bank when the input signal is reconstructed from a reduced number of channels. This criterion mirrors the principal component approximations of the Karhunen-Loève transform. An optimum compaction filter bank is constructed as follows: an optimum compaction filter is designed first and a filter bank procedure renders the remaining filters.

The constraint that the optimum compaction filter be Nyquist(M) is

$$|H(e^{i\omega})|^2_{\downarrow M} = 1$$

or, equivalently

$$\sum_{k=0}^{M-1} |H(e^{i(\omega-2k\pi/M)})|^2 = M, \qquad \omega \in [0, 2\pi]$$

is imposed because the analysis filters of an orthonormal filter bank always satisfy it.

Theorem 3.4.1. *(Construction of an optimum compaction filter.) Assume that the input power spectral density is piecewise continuous on $[0, 2\pi]$. The following procedure renders an optimum compaction filter:*

- *Step 1. At each frequency ω_0 in $[0, 2\pi/M)$ define the M alias frequencies by*

$$\omega_k = \omega_0 + 2k\pi/M, \qquad 0 \le k \le M - 1\,.$$

- *Step 2. Compare the values of $S_{xx}(e^{i\omega_k})$ at these M alias frequencies $\{\omega_k\}_{0 \le k \le M-1}$. Let L be the smallest integer such that $S_{xx}(e^{i\omega_L})$ is a maximum in this set. Assign*

$$H(e^{i(\omega_0 + 2k\pi/M)}) = \begin{cases} \sqrt{M}, & \text{if } k = L, \\ 0, & \text{otherwise.} \end{cases}$$

- *Repeat for each ω_0 in $[0, 2\pi/M)$. $H(e^{i\omega})$ is defined for all w in $[0, 2\pi]$ and satisfies the conditions from Definition 3.4.1.*

Proof. [118] Evaluate the output variance taking into account the construction of the filter H. First, the output power spectral density is

$$\begin{aligned} S_{vv}(e^{i\omega}) &= S_{xx}(e^{i\omega})|H(e^{i\omega})|^2 \\ &= \frac{1}{M} \sum_{k=0}^{M-1} S_{xx}(e^{i(\omega + 2\pi k/M)})|H(e^{i(\omega + 2\pi k/M)})|^2\,. \end{aligned}$$

86
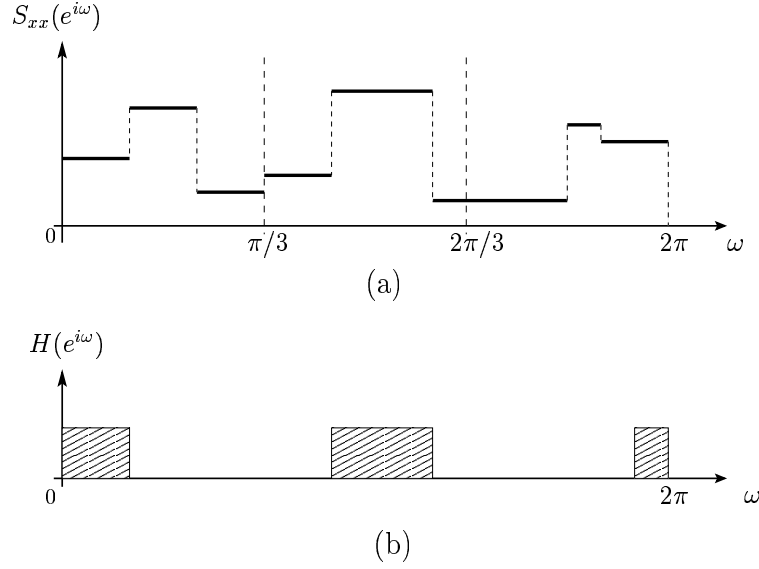
Figure 3.7: (a) Input power spectral density. (b) Optimal compaction filter in a 3-channel filter bank.

The output variance is

$$
\begin{aligned}
\sigma_v^2 &= \int_0^{2\pi} S_{vv}(e^{i\omega})d\omega/(2\pi) \\
&= \frac{1}{M}\int_0^{2\pi}\sum_{k=0}^{M-1} S_{xx}(e^{i(\omega+2\pi k/M)})|H(e^{i(\omega+2\pi k/M)})|^2 d\omega/(2\pi). \quad (3.19)
\end{aligned}
$$

The integrand should be maximized for each ω in $[0, 2\pi)$ by an appropriate choice of $|H(\cdot)|^2$. Let $\omega_0 \in [0, 2\pi/M)$. Denote by L the smallest integer such that

$$
S_{xx}(e^{i(\omega+2\pi L/M)}) = \max_k\{S_{xx}(e^{i(\omega+2\pi k/M)})\},
$$

and recall the Nyquist(M) constraint from Equation (3.10):

$$
\sum_{l=0}^{M-1} |H_k(e^{i(\omega-2\pi l/M)})|^2 = M, \qquad 0 \le k < M.
$$

87

Consider a sum $S = \sum_{k=0}^{M-1} c_k x_k$ where c_k are fixed nonnegative numbers such that $c_0 \geq c_1 \geq \ldots \geq c_{M-1} \geq 0$ and x_k are sought such that the sum S is maximum under the constraint $\sum_{k=0}^{M-1} x_k = x > 0$. Then the best choice of the constants x_k is given by $x_0 = x$ and $x_k = 0$, $k \neq 0$.

Thus the integrand in Equation (3.19) is maximized for a fixed frequency w_0 under the Nyquist(M) constraint if

$$H(e^{i(\omega_0 + 2\pi k/M)}) = \begin{cases} M, & k = L \\ 0, & k \neq L, \ \ 0 \leq k < M \end{cases}$$

This choice satisfies the Nyquist(M) constraint. A filter is constructed with passband width of $2\pi/M$ and its output variance is the largest among all the Nyquist filters with the same passband width. $\qquad\square$

An example of this procedure is shown in Figure 3.7.

3.4.1.1 Remarks on the construction of optimum compaction filters

- The construction of the previous theorem renders ideal passband filters: $H(\cdot)$ is an ideal two-level filter with passband response $\sqrt{M}$ and stopband response 0.

- The magnitude $|H(\cdot)|$ is not uniquely determined because of the selection(choice) strategy at the comparison step in Theorem 3.4.1.

- If $H(\cdot)$ is an optimal compaction filter for a given input power spectral density $S_{xx}(\cdot)$ then it will be a valid optimal solution for any transformed $f(S_{xx}(\cdot))$ as long as $f(.)$ is a positive nondecreasing function.

88

- For real inputs, the construction of $H(\cdot)$ can be modified to obtain a symmetric solution.

- If the input is white, any filter satisfying the Nyquist(M) condition is optimal.

- Filters constructed according to this algorithm are unrealizable ideal infinite order filters. Approximations of these filters with finite impulse response filters are used in the performance analysis to compare with the coding gain of realizable filters [115].

3.4.2 Energy compaction eigenproblem

The next result shows the correspondence between optimum compaction gain and optimum coding gain in $\mathcal{U}_{\infty,M}$. Consider the polyphase representation of an analysis filter:

$$H(z) = \sum_{k=0}^{M-1} z^{-k} E_k(z^M).$$

The Nyquist(M) condition can be expressed equivalently as:

$$\sum_{k=0}^{M-1} |E_k(e^{i\omega})|^2 = 1, \qquad \omega \in [0, 2\pi).$$

Define:

$$\boldsymbol{e}^\dagger(e^{i\omega}) := [E_0(e^{i\omega}) \ E_1(e^{i\omega}) \ ... \ E_{M-1}(e^{i\omega})].$$

The Nyquist(M) condition becomes:

$$\boldsymbol{e}^\dagger(e^{i\omega})\boldsymbol{e}(e^{i\omega}) = 1, \qquad \omega \in [0, 2\pi].$$

89

Denote by $\boldsymbol{S_{xx}}(\cdot)$ the power spectral density matrix of the blocked process $\boldsymbol{x} : \mathbb{Z} \to \mathbb{C}^M$. The power spectral density of the output $v : \mathbb{Z} \to \mathbb{C}$ is given by:

$$S_{vv}(e^{i\omega}) = \boldsymbol{e}^\dagger(e^{i\omega})\boldsymbol{S_{xx}}(e^{i\omega})\boldsymbol{e}(e^{i\omega})$$

and the variance:

$$\sigma_v^2 = \int_0^{2\pi} \boldsymbol{e}^\dagger(e^{i\omega})\boldsymbol{S_{xx}}(e^{i\omega})\boldsymbol{e}(e^{i\omega})\frac{d\omega}{2\pi} .$$

Maximization of σ_v^2 under the Nyquist(M) constraint is equivalent to the maximization of the integrand; thus $\boldsymbol{e}(e^{i\omega})$ is the eigenvector corresponding to the maximum eigenvalue of $\boldsymbol{S_{xx}}(e^{i\omega})$ at any value of the argument ω. The above steps constitute the proof of the next theorem.

Theorem 3.4.2. [113, 118] *A filter is an optimum compaction filter for a signal with given input power spectral density if and only if the following conditions are satisfied:*

1. *the polyphase vector of the filter is para-unitary, or equivalently the filter satisfies the Nyquist(M) constraint and*

2. *the polyphase vector of the filter is an eigenvector corresponding to the largest eigenvalue of the input blocked power spectral density matrix.*

3.5   Design Considerations

3.5.1   Polyphase interpretation of optimality

Here it is shown the connection between the two concepts of optimality: in the coding gain sense and with respect to the maximum compaction gain.

Theorem 3.5.1. *A coding gain optimal filter bank over* $\mathcal{U}_{\infty,M}$ *contains an optimum compaction filter.*

Proof. [118] Assume that the output sub-band variances ordered decreasingly; we show that the first filter in the bank, $H_0(\cdot)$ is an optimum compaction filter.

From the Theorem 3.3.3 on necessary and sufficient conditions for optimality we know that the polyphase matrix $\boldsymbol{E}(\cdot)$ of a coding gain optimal filter bank in $\mathcal{U}_{\infty,M}$ diagonalizes the blocked power spectral density matrix.

$$\boldsymbol{E}(e^{i\omega})\boldsymbol{S_{xx}}(e^{i\omega})\boldsymbol{E}^{\dagger}(e^{i\omega}) = \mathrm{diag}\{S_0(e^{i\omega}),\ S_1(e^{i\omega}),...,S_{M-1}(e^{i\omega})\}$$

Denote as in Theorem 3.4.2 by $\boldsymbol{e}_k$ the eigenvectors of the input power spectral density matrix. The relation to the polyphase matrix of an optimal coding gain filter bank is

$$\begin{bmatrix} \boldsymbol{e}_0^{\dagger}(e^{i\omega}) \\ \boldsymbol{e}_1^{\dagger}(e^{i\omega}) \\ \vdots \\ \boldsymbol{e}_{M-1}^{\dagger}(e^{i\omega}) \end{bmatrix} \quad \boldsymbol{S_{xx}}(e^{i\omega}) \quad [\boldsymbol{e}_0(e^{i\omega})\ \boldsymbol{e}_1(e^{i\omega})\ \cdots\ \boldsymbol{e}_{M-1}(e^{i\omega})]$$

$$= \begin{bmatrix} S_0(e^{i\omega}) & 0 & \cdots & 0 \\ 0 & S_1(e^{i\omega}) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & S_{M-1}(e^{i\omega}) \end{bmatrix} \tag{3.20}$$

where we identify:

$$\boldsymbol{E}(e^{i\omega}) = \begin{bmatrix} \boldsymbol{e}_0^{\dagger}(e^{i\omega}) \\ \boldsymbol{e}_1^{\dagger}(e^{i\omega}) \\ \vdots \\ \boldsymbol{e}_{M-1}^{\dagger}(e^{i\omega}) \end{bmatrix}.$$

91

Since $\boldsymbol{E}(e^{i\omega})$ is unitary for all ω, the sub-band spectra

$S_k(e^{i\omega})$, $k = 0, 1, \ldots, M-1$ are the eigenvalues and the columns of $\boldsymbol{E}(e^{i\omega})$ form

an orthonormal set of corresponding eigenvectors of $\boldsymbol{S_{xx}}(e^{i\omega})$. The spectral

majorization property is a necessary condition for optimality therefore $S_0(e^{i\omega})$

is the largest eigenvalue of $\boldsymbol{S_{xx}}(e^{i\omega})$ and $\boldsymbol{e}_0(e^{i\omega})$ is the corresponding eigen-

vector, for each ω. Theorem 3.4.2 is invoked to complete the proof. □

3.5.2 Procedure to find optimal analysis filter banks

An optimal filter bank can be designed based on the result in the pre-

vious theorem [118, 63]. One optimal compaction filter is constructed for the

given input power spectral density followed by a sequence of optimal com-

paction filters for the frequencies that are not part of the previously con-

structed filters. The input x is a wide-sense cyclo-stationary process with a

zero mean and given power spectral density $S_{xx}(\cdot)$. The noise is modeled with

the standard model as described in Section 1.4. The desired filter bank is also

optimal with respect to the coding gain in $\mathcal{U}_{\infty,M}$.

The steps described next are called the "peel off" procedure [118]. Fig-

ure 3.8 illustrates the procedure. Suppose $H_0(\cdot)$ is constructed as an optimal

energy compaction filter with passband support Ω_0 as in Theorem 3.4.1. Define

a partial power spectral density function by

$$S_{xx}^{(1)}(e^{i\omega}) = \begin{cases} 0, & \omega \in \Omega_0 \\ S_{xx}(e^{i\omega}), & \text{otherwise}. \end{cases}$$

The next filter $H_1(\cdot)$ is designed as the optimal compaction filter for the partial
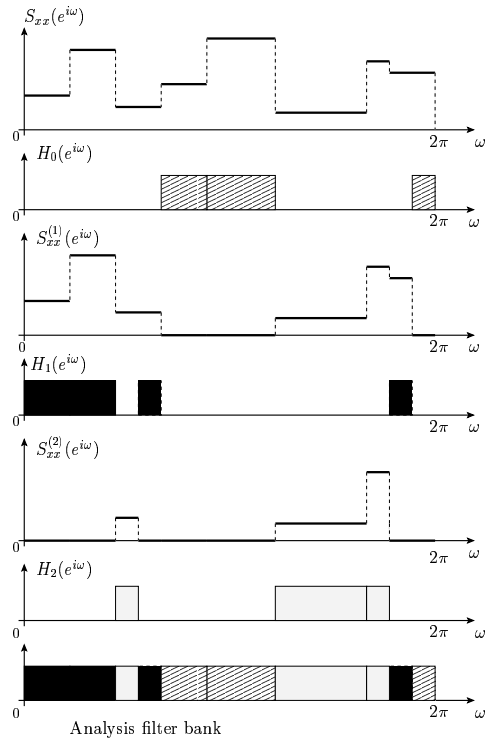
Figure 3.8: Design of ideal signal-adapted filter banks

power spectral density $S_{xx}^{(1)}(e^{i\omega})$ and denote by Ω_1 its passband. Define

$$S_{xx}^{(2)}(e^{i\omega}) = \begin{cases} 0, & \omega \in \Omega_0 \cup \Omega_1 \\ S_{xx}(e^{i\omega}), & \text{otherwise}. \end{cases}$$

and denote by Ω_1 its passband The procedure continues until the last filter

93

$H_{M-1}(\cdot)$ is constructed. The filters designed using the "peel-off" procedure realize a total decorrelation because they do not overlap. These are *ideal unrealizable* filters. [115, 114].

Remark 3.5.1. The output has the spectral majorization property.

Proof. Let ω_0 be arbitrary from the interval $[0, 2\pi/M)$ and define

$$\omega_k = \omega_0 + 2\pi k/M, \qquad k = 0, 1, \ldots, M - 1.$$

The w_k's are the alias frequencies of ω_0. By construction, each filter H_k has exactly one of these frequencies in its passband. The compaction filters are designed in a sequential order, the first ones corresponding to higher values of the input power spectral density. The value of the square frequency response of each filter is constant, $\sqrt{M}$. Renumber the ω_k's such that ω_l is in the passband of H_l. Then

$$S_{xx}(e^{i\omega_0}) \geq S_{xx}(e^{i\omega_1}) \geq \ldots \geq S_{xx}(e^{i\omega_{M-1}}), \qquad \omega \in [0, 2\pi).$$

The down-sampled power spectral density $S_l(\cdot)$ takes the value of $S_{xx}(e^{i\omega_l})$ at the frequency $M\omega_l$(modulo 2π). Denote

$$\widehat{\omega}_l = M\omega_l(\text{modulo } 2\pi).$$

The role of the modulo operation is to keep the frequencies within $[0, 2\pi])$, hence $\widehat{\omega}_l \in [0, 2\pi]$. Since the ω_l's are alias frequencies the quantity

$$\widehat{\omega} = M\omega_l(\text{modulo } 2\pi)$$

is the same for all $l = 0, 1, \ldots, M - 1$. Hence from the previous inequality it follows that

$$S_0(e^{i\widehat{\omega}}) \geq S_1(e^{i\widehat{\omega}}) \geq \ldots \geq S_{M-1}(e^{i\widehat{\omega}})$$

holds for all $\hat{\omega}$ in $[0, 2\pi)$. Since ω_0 is chose arbitrarily in $[0, 2\pi/M)$ the inequality holds for all $\widehat{\omega} \in [0, 2\pi)$. Hence, the majorization of sub-band spectra holds.

$\square$

The filter banks designed through the above procedure are both coding gain and compaction gain optimal. These filter banks are fully signal-adapted; however they are ideal filters. Their practical use is in performance evaluation: the ideal filters are approximated by finite impulse response filter banks and the estimated coding gain of an ideal filter bank is compared against the coding gain of a given filter bank.

The compaction optimality is a desired property in certain applications such as compression, signal analysis and there are strategies to construct finite impulse response optimum compaction filter banks that are signal-adaptive. For examples, see [80, 79, 63]. However, a construction of finite impulse response filter banks following the "peel-off" procedure is a seemingly intractable nonlinear optimization problem for which no solution is known and no algorithm has yet been found satisfactory. See [80, 114, 113]. In the next chapter the compaction and coding gain issues and design of finite impulse response filter banks are discussed in the context of principal component optimality.

Chapter 4

Principal Component Filter Banks

4.1 Introduction

This chapter presents a new approach to optimality of filter banks in a sense that couples the Karhunen-Loève transform optimality with the coding gain and the compaction gain optimalities.

The mathematical setting for this new approach is introduced by Unser [114] who raises the question of existence and design of a para-unitary M-channel filter bank optimal in the sense of minimizing the reconstruction error when fewer than M sub-bands are used for signal reconstruction. This is a direct extension of the optimality exhibited by the Karhunen-Loève transform for the class of transform coders.

The major drawback of the Karhunen-Loève transform is its signal dependence which makes it very expensive. In practice, suboptimal transforms and filter banks are used, such as discrete cosine transform, lapped orthogonal transforms and wavelet transforms [109, 89, 36, 29]. One reasonable approach is to design filter banks whose performance parallels that of the Karhunen-Loève transform, at a lower cost if possible. In order to pursue this path, we first study in depth the existence of filter banks that couple the principal

component optimality with the coding gain maximization.

A first challenge posed by the design of a filter bank is the length of the filters. Finite impulse response filters are of practical importance, of course; however the theory is developed on the universe of orthonormal filter banks $\mathcal{U}_{\infty,M}$ without restriction on the filter orders. The strategy is to develop theory on the unrestricted class of filters and to apply it to finite impulse response filters, if possible.

The concept of a principal component filter bank as a perfect reconstruction one which minimizes the approximation error – in the mean-square sense – between the original signal and its low-resolution version was introduced by Tsatsanis and Giannakis [113]. A filter design is sought such that its first P sub-bands approximate best the input signal in the mean-square sense. Let a zero-mean wide sense stationary discrete stochastic process be input to a uniform maximally down-sampled M-channel para-unitary filter bank. Suppose that the output sub-band variances are ordered decreasingly and that we use only the first P sub-bands to reconstruct the signal. The approximation error is the average of the discarded sub-band variances

$$\epsilon(P) = \frac{1}{M} \sum_{k=P}^{M-1} \sigma_{v_k}^2, \qquad P \leq M \tag{4.1}$$

with the variances given by Equation (3.12):

$$\sigma_{v_k}^2 = \int_0^{2\pi} \left[\boldsymbol{E}(e^{i\omega}) \boldsymbol{S_{vv}}(e^{i\omega}) \boldsymbol{E}^\dagger(e^{i\omega}) \right]_{kk} d\omega/(2\pi), \qquad k = 0, 1, \ldots, M-1.$$

The input signal is decomposed into *uncorrelated* low-resolution principal components with decreasing variance. This technique for filter bank design has

97

its origin in the results for principal components for time-series analysis [17]. The mean square error criterion is commonly used in signal processing; nevertheless many results on the optimality of principal component filter banks [130, 131] emerged under other optimality criteria [80, 79, 63, 8, 7].

This chapter is structured as follows. The first results on principal component filter banks as they are derived from time-series analysis are shown. A rigorous definition of principal component filter banks is provided. Two sections show our contributions: a new theorem on the existence of coding gain optimal filter banks for finite impulse response filter banks and a study of principal component filter banks and optimal coding gain for filter banks with McMillan degree 1. Finally, the existence of principal component filter banks is discussed.

4.2   Principal Component Filter Banks from Time-Series Analysis

The first results on principal component filter banks are derived by Tsatsanis and Giannakis [113].

Definition 4.2.1. A process is *second-order* if its mean and autocorrelation are ergodic, i.e. statistical averages are estimated as time averages:

$$E\{x\} \;=\; \lim_{N\to\infty} \frac{1}{2N+1} \sum_{k=-N}^{N} x[n]\,,$$

$$R_{xx}[k] \;=\; \lim_{N\to\infty} \frac{1}{2N+1} \sum_{k=-N}^{N} x[n]\,x^{\dagger}[n-k]\,.$$

98

Let $\boldsymbol{x}$ be a M-dimensional zero-mean second-order real valued process and assume that $\widehat{\boldsymbol{x}}$ is an approximation of $\boldsymbol{x}$. The time-averaged mean square reconstruction error is defined as

$$J = \lim_{N \to \infty} \frac{1}{2N+1} \sum_{n=-N}^{N} E\{(\boldsymbol{x}[n] - \widehat{\boldsymbol{x}}[n])^T (\boldsymbol{x}[n] - \widehat{\boldsymbol{x}}[n])\}$$

Let the autocorrelation matrix

$$\boldsymbol{R_{xx}}[n, \tau] = E\{\boldsymbol{x}[n+\tau]\boldsymbol{x}^T[n]\}$$

satisfy a uniform boundedness criterion for all integers n:

$$|[\boldsymbol{R_{xx}}[n, \tau]]_{k,l}| \le B(\tau), \quad \sum_{\tau} B(\tau) < \infty \qquad 1 \le k, l \le M. \qquad (4.2)$$

Define the time-averaged autocorrelation and power spectral density matrices, $\overline{\boldsymbol{R_{xx}}}[\tau]$ and $\overline{\boldsymbol{S_{xx}}}(\omega)$, respectively, as

$$\overline{\boldsymbol{R_{xx}}}[\tau] = \lim_{N \to \infty} \frac{1}{2N+1} \sum_{k=-N}^{N} \boldsymbol{R_{xx}}[n, \tau] \text{ and}$$

$$\overline{\boldsymbol{S_{xx}}}(\omega) = \sum_{\tau \in \mathbb{Z}} \overline{\boldsymbol{R_{xx}}}[\tau] e^{-i\omega\tau}.$$

Refer to Figure 3.2. Denote by $\boldsymbol{E}(\cdot)$ and $\boldsymbol{R}(\cdot)$ the matrices whose rows are the M-polyphase components of each filter from the original analysis filter bank $H_0(\cdot), H_1(\cdot), \ldots, H_{M-1}(\cdot)$ and synthesis filter bank $F_0(\cdot), F_1(\cdot), \ldots, F_{M-1}(\cdot)$ in Figure 3.1. The matrices $\boldsymbol{E}(\cdot)$ and $\boldsymbol{R}(\cdot)$ are sought such that the filter bank is optimal in the sense of minimizing the mean-square projection error – see Section 2.3. In the blocked representation of the input and of the sub-band

signals, the action of the partial filters is

$$\boldsymbol{v}[n] \;=\; \sum_{k \in \mathbb{Z}} \boldsymbol{E}_{n-k}\boldsymbol{x}[k]\,,$$

$$\widehat{\boldsymbol{x}}[n] \;=\; \sum_{k \in \mathbb{Z}} \boldsymbol{R}_{n-k}\boldsymbol{v}[k], \qquad n, k \in \mathbb{Z}\,,$$

where it is understood that the polyphase matrices are expressed as:

$$\boldsymbol{E}(z) \;=\; \sum_{n \in \mathbb{Z}} \boldsymbol{E}_n z^{-n}$$

$$\boldsymbol{R}(z) \;=\; \sum_{n \in \mathbb{Z}} \boldsymbol{R}_n z^{-n}$$

and this expression is valid in $\mathcal{U}_{\infty,M}$. If only a finite number of impulse response matrices $\boldsymbol{E}_n$, with positive indices, are non-zero then the polyphase matrices correspond to a finite impulse response filter bank with causal analysis filters

$$\boldsymbol{E}(z) = \sum_{k=0}^{K} \boldsymbol{E}_k z^{-k}$$

The next proposition addresses the question of design of these polyphase matrices that are optimal in the principal component sense.

Proposition 4.2.1. [113]. *Let $\boldsymbol{x}$ be M-dimensional zero-mean, second-order real valued process satisfying the uniform boundedness criterion from Equation (4.2). Then the optimal $P \times M$ and $M \times P$ polyphase matrices $\boldsymbol{E}(\cdot), \boldsymbol{R}(\cdot)$ that minimize the time-averaged mean-square error are*

$$\boldsymbol{E}(e^{i\omega}) \;=\; \begin{bmatrix} \boldsymbol{e}_0^{\dagger}(e^{i\omega}) \\ \boldsymbol{e}_1^{\dagger}(e^{i\omega}) \\ \vdots \\ \boldsymbol{e}_{P-1}^{\dagger}(e^{i\omega}) \end{bmatrix}$$

$$\boldsymbol{R}(e^{i\omega}) \;=\; \boldsymbol{E}^{\dagger}(e^{i\omega}), \qquad \omega \in [0, 2\pi]$$

100

where $\boldsymbol{e}_k(e^{i\omega})$, $k = 0, 1, \ldots, P - 1$ is the eigenvector corresponding to the $(k + 1)$-st largest eigenvalue of the time-averaged input power spectral density matrix at each ω in $[0, 2\pi]$.

The next corollary describes the filters in the banks for a scalar process.

Corollary 4.2.2. [113].   *Let x be a second-order process with uniformly bounded autocorrelation. Then the optimal filters $H_k(\cdot), F_k(\cdot), k = 0, 1, ..., P-1$ that minimize the time-averaged mean-square reconstruction error*

$$J = \lim_{N \to \infty} \frac{1}{2N + 1} \sum_{n=-N}^{N} E\{[\hat{x}[n] - x[n]]^2\}$$

are given by

$$
\begin{aligned}
H_k(z) &= \sum_{l=0}^{M-1} z^{-l} E_{kl}(z^M) \\
F_k(z) &= H_k(z^{-1}), \qquad k = 0, 1, \ldots, P - 1
\end{aligned}
\tag{4.3}
$$

where $z = e^{i\omega}$, $\omega \in [0, 2\pi]$ and $\boldsymbol{e}_k^\dagger(e^{i\omega}) = [E_{k0}(e^{i\omega}) , ..., E_{k,M-1}(e^{i\omega})]$ is the eigenvector corresponding to the $(k + 1)st$ largest eigenvalue of the time-averaged input power spectral density matrix.

Note the similarity between the assertion in this corollary and the Theorem 3.4.2 and Equation (3.20). The corollary states that the polyphase matrices derived from principal components render the least mean-square reconstruction error when only a few sub-bands are used to reconstruct the signal. The next two propositions assure that these filters and polyphase matrices form an orthonormal filter bank when all the principal components are used $(P = M)$.

Proposition 4.2.3. [113]. *The polyphase matrices constructed as in Proposition 4.2.1 and the corresponding filters from Equation (4.3) of Corollary 4.2.2 constitute a perfect reconstruction filter bank when all the sub-bands are considered for the reconstruction of the input signal.*

The proof is straightforward by showing that the resulting polyphase matrices are para-unitary; and that is exactly what Equation (3.20) flags. Now a result similar to the Theorem 3.3.3 on necessary and sufficient conditions for optimality:

Proposition 4.2.4. [113]. *If a M-channel filter bank is designed according to Proposition 4.2.1 and its corollary, then the sub-band signals are uncorrelated with each other the time-averaged input power spectral density matrix is diagonalized by such a filter bank.*

The detailed proofs of all these results can be found in [113], where similar results are also derived for the familiar wide sense stationary input. These results for the second-order real-valued processes with time-averaged bounded auto-correlation show how the optimality approach in the principal component sense emerged from Brillinger's time-series analysis [17].

4.3 Definition of Principal Component Filter Banks

Consider a para-unitary filter bank in $\mathcal{U}_{\infty,M}$ (Figure 3.3). The input is a wide sense cyclo-stationary discrete stochastic process with zero mean. The given input power spectral density $S_{xx}(\cdot)$ may vanish at most on a set

of measure zero. A principal component filter bank in $\mathcal{U}_{\infty,M}$ is sought such that it minimizes the reconstruction error when the signal is approximated by its first principal components. The expression for the reconstruction error is given by Equation (4.1)

$$\epsilon(P) = \frac{1}{M} \sum_{k=P}^{M-1} \sigma_{v_k}^2, \qquad P \le M$$

with the convention that the output sub-band variances are ordered decreasingly. The minimization of this error amounts to maximization of the sum of the variances of the sub-bands used to reconstruct the signal

$$S(P) = \frac{1}{M} \sum_{k=0}^{P-1} \sigma_{v_k}^2, \qquad P \le M \,.$$

It is desirable to link the principal component optimality with the coding gain, since the coding gain is such a widely used metric. In other words it is desirable to answer the question: " If a filter bank is such that the sum $S(\cdot)$ is maximized in $\mathcal{U}_{\infty,M}$ for any value of P in $\{0, 1, \ldots, M-1\}$, is the coding gain of that filter bank also optimal in $\mathcal{U}_{\infty,M}$?". Or even better, for practical purposes, the same question should be rephrased for classes $\mathcal{C}$ of filter banks that are proper subsets of $\mathcal{U}_{\infty,M}$, like causal finite impulse response filter banks.

The issue of comparing coding gain optimality with maximization of $S(\cdot)$ is investigated first in $\mathcal{U}_{\infty,M}$. A few preliminary theoretical results are necessary and majorization theory is employed. See [15, 47, 48, 73, 9]. The definition of the majorization relation and terminology was introduced by Hardy,

103

Littlewood and Polya in [46] and the theory of majorization is studied exten-
sively by Marshall and Olkin in [73]. Many valuable applications to matrix
analysis can be found in the books by Horn and Johnson in [48, 47].

Definition 4.3.1. Consider the vectors $\boldsymbol{\sigma}, \boldsymbol{\tau} \in \mathbb{R}^M$, $\boldsymbol{\sigma} = [\sigma_0, \sigma_1, \ldots, \sigma_{M-1}]^T$,
$\boldsymbol{\tau} = [\tau_0, \tau_1, \ldots, \tau_{M-1}]^T$ with the components ordered decreasingly

$$\sigma_0 \geq \sigma_1 \geq \ldots \geq \sigma_{M-1}, \quad \tau_0 \geq \tau_1 \geq \ldots \geq \tau_{M-1}.$$

Define:

$$\boldsymbol{\sigma} \succ \boldsymbol{\tau} \text{ if } \begin{cases} \sum_{k=0}^{P} \sigma_k \geq \sum_{k=0}^{P} \tau_k, & 0 \leq P < M-1 \\ \sum_{k=0}^{M-1} \sigma_k = \sum_{k=0}^{M-1} \tau_k \end{cases}.$$

When $\boldsymbol{\sigma} \succ \boldsymbol{\tau}$ it is said that $\boldsymbol{\sigma}$ *majorizes* $\boldsymbol{\tau}$ (or $\boldsymbol{\tau}$ is majorized by $\boldsymbol{\sigma}$).

We may also view the entries of a vector as the elements of a set (with
some elements duplicated) and think of the majorization defined on sets as
well. The ordering of the components is a simplifying hypothesis for the use
of majorization in filter bank theory, since we always consider that the output
sub-band variances are ordered decreasingly.

Definition 4.3.2. A real valued function ϕ defined on a set $\mathcal{S} \subset \mathbb{R}^M$ is said
to be *Schur-convex* on $\mathcal{S}$ if

$$\boldsymbol{\sigma} \succ \boldsymbol{\tau} \Longrightarrow \phi(\boldsymbol{\sigma}) \geq \phi(\boldsymbol{\tau})$$

for all $\boldsymbol{\sigma}$, $\boldsymbol{\tau}$ in $\mathcal{S}$. If $\boldsymbol{\sigma} \succ \boldsymbol{\tau} \Longrightarrow \phi(\boldsymbol{\sigma}) \leq \phi(\boldsymbol{\tau})$ then ϕ is said to be *Schur-concave*.

The next proposition connects the ordinary convex functions with the
majorization relation. See [73].

Proposition 4.3.1. *(Schur). If $I \subset \mathbb{R}$ is an interval and $g : I \to \mathbb{R}$ is convex then the function ϕ defined as $\phi(\boldsymbol{\sigma}) = \sum_{k=0}^{M-1} g(\sigma_k)$, $\boldsymbol{\sigma} = [\sigma_0, \sigma_1, \ldots, \sigma_{M-1}]^T$ is Schur-convex on I^M.*

These results are now applied to the theory of orthonormal filter banks. Assume that the input variance is normalized: $\sigma_x^2 = 1$. Recall that the sum of the output sub-band variances is constant

$$\sum_{k=0}^{M-1} \sigma_{v_k}^2 = M$$

as shown in Proposition 3.2.1. Each sub-band variance vector has only strictly positive entries by Proposition 3.2.2, because we assume that the power spectral density may vanish only on sets of measure zero. Note that this assumption ensures that the coding gain is a well defined function of the sub-band variances for all filter banks in $\mathcal{U}_{\infty,M}$. We also have the bound on the sub-band variances

$$0 < \sigma_{v_k}^2 < 1, \qquad k = 0, 1, \ldots, M - 1 .$$

The power spectral density matrix of the wide sense stationary blocked input is symmetric and positive semidefinite [38, 37]. An output sub-band variance vector that majorizes all other sub-band variance vectors that are output from filter banks in $\mathcal{U}_{\infty,M}$ corresponds to a filter bank that is optimal in the mean-square projection error sense, and we call it a principal component filter bank. Note that this optimality is over the unconstrained universe $\mathcal{U}_{\infty,M}$. A formal

105

definition of principal component filter banks for arbitrary classes in $\mathcal{U}_{\infty,M}$ will be given after some preliminary theory.

We now connect this result to the coding gain. Denote by $\boldsymbol{\sigma}$ the output sub-band variance vector $\boldsymbol{\sigma} = [\sigma_0, \sigma_1, \ldots, \sigma_{M-1}]^T$. From the arithmetic/geometric mean inequality we conclude that the coding gain is greater than one:

$$
\begin{aligned}
G_{coding} \quad &: \quad (0,1)^M \to [1, \infty) \,, \\
G_{coding}(\boldsymbol{\sigma}) \quad &= \quad \frac{1}{(\prod_{k=0}^{M-1} \sigma_{v_k}^2)^{1/M}} \,.
\end{aligned}
$$

The minimum value of the coding gain is attained when all the sub-band variances are equal. Coding gain optimality, i.e., maximization of the coding gain over $\mathcal{U}_{\infty,M}$, is equivalent to minimization of the product of the output sub-band variances. We investigate when this product attains its minimum. Consider the convex function

$$
\begin{aligned}
g \quad &: \quad (0, \infty) \to \mathbb{R} \,, \\
g(\sigma) \quad &= \quad -\log(\sigma) \,.
\end{aligned}
$$

We now write the function g of components of the sub-band variance vector, $\boldsymbol{\sigma} = [\sigma_{v_0}^2, \sigma_{v_1}^2, \ldots \sigma_{v_{M-1}}^2]^T$, sum up and obtain

$$
\begin{aligned}
\sum_{k=0}^{M-1}(-\log(\sigma_{v_k}^2)) \quad &= \quad M[\frac{1}{M}\sum_{k=0}^{M-1}(-\log(\sigma_{v_k}^2))] = \\
&= \quad M\log\frac{1}{(\prod_{k=0}^{M-1}\sigma_{v_k}^2)^{1/M}} \\
&= \quad M\log G_{coding}(\boldsymbol{\sigma}) \,.
\end{aligned}
$$

106

By Proposition 4.3.1 the function $\phi(\boldsymbol{\sigma}) = \log G_{coding}(\boldsymbol{\sigma})$ is a Schur-convex function. The logarithm is a strictly increasing function, hence the inequality $\log(G_{coding}(\boldsymbol{\sigma})) \geq \log(G_{coding}(\boldsymbol{\tau}))$ is equivalent to $G_{coding}(\boldsymbol{\sigma}) \geq G_{coding}(\boldsymbol{\tau})$ on $(0,1)^M$. We conclude that

$$\boldsymbol{\sigma} \succ \boldsymbol{\tau} \Longrightarrow G_{coding}(\boldsymbol{\sigma}) \geq G_{coding}(\boldsymbol{\tau}) \,.$$

Remark 4.3.1. The previous result is a equivalent to saying that if a vector $\boldsymbol{\sigma}$ majorizes a vector $\boldsymbol{\tau}$ then the product of the entries of $\boldsymbol{\sigma}$ is less than the product of the entries of $\boldsymbol{\tau}$ [46].

The majorization of output sub-band variance vector and coding gain optimality over $\mathcal{U}_{\infty,M}$ are therefore equivalent. Over the whole universe $\mathcal{U}_{\infty,M}$ the optimum compaction gain and the optimum coding gain occur simultaneously, for the same filter bank – which is an ideal filter bank according to the results from the previous chapter. The compaction gain optimality is understood in the sense of the construction from Section 3.5.2, the "peel-off" procedure.

The results on coding gain and principal component optimalities are coupled by the study of the output sub-band variance vector from the majorization perspective. Next the sub-band spectra are discussed.

Recall that Theorem 3.3.3 states that a necessary condition for coding gain optimality is that the sub-band spectra satisfy the spectral majorization as in Definition 3.2.2. Hence we investigate the equivalence of the next two

inequalities:

$$\sigma_{v_0}^2 \geq \sigma_{v_1}^2 \geq ... \geq \sigma_{v_{M-1}}^2 , \tag{4.4}$$

$$S_0(e^{i\omega}) \geq S_1(e^{i\omega}) \geq ... \geq S_{M-1}(e^{i\omega}), \qquad \omega \in [0, 2\pi] . \tag{4.5}$$

The output sub-band variances are given by

$$\sigma_{v_k}^2 = \int_0^{2\pi} S_k(e^{i\omega}) d\omega / (2\pi) .$$

Obviously the spectral majorization implies sub-band variance ordering – in the sense of Equation (4.4). It turns out that for an optimal filter bank in $\mathcal{U}_{\infty,M}$ the converse also holds – see the proof of Theorem 3.3.2.

Remark 4.3.2. The equivalence between majorization of sub-band spectra and that of the sub-band variances is valid only for the filter bank whose coding gain cannot be increased in $\mathcal{U}_{\infty,M}$. That filter bank is the one that diagonalizes the input power spectral density matrix.

The vector of the diagonal elements of the output power spectral density matrix is majorized by the vector of its eigenvalues at each frequency. By Hadamard's Inequality the product of the diagonal elements of the output power spectral density matrix is greater than its determinant, which is the product of its eigenvalues. Denote the output power spectral density matrix by $\boldsymbol{S_{vv}}(\cdot)$ and denote by $\lambda_k(\cdot)$ its eigenvalues, calculated at each frequency. For each $\omega \in [0, 2\pi]$ denote the output power spectral density matrix elements by $S_{kl}(e^{i\omega})$, $k, l = 0, 1, \ldots, M-1$. then the vector of diagonal elements is:

$$\boldsymbol{s}(e^{i\omega}) = [S_{00}(e^{i\omega}), S_{11}(e^{i\omega}), \ldots, S_{M-1,M-1}(e^{i\omega})]^T .$$

Let $\boldsymbol{\lambda}(\cdot)$ be the eigenvector of $\boldsymbol{S_{vv}}(\cdot)$:

$$\boldsymbol{\lambda}(e^{i\omega}) = [\lambda_0(e^{i\omega}), \lambda_1(e^{i\omega}), \ldots, \lambda_{M-1}(e^{i\omega})]^T.$$

Hadamard's Inequality and the definition of majorization give, at each $\omega \in [0, 2\pi]$:

$$\prod_{k=0}^{M-1} \lambda_k(e^{i\omega}) \leq \prod_{k=0}^{M-1} S_{kk}(e^{i\omega}) \text{ and}$$
$$\boldsymbol{\lambda}(e^{i\omega}) \succ \boldsymbol{s}(e^{i\omega}).$$

The last inequalities are the ones from Theorem 3.3.3. When the output power spectral density matrix is diagonal, the vector of its eigenvalues majorizes any vector of diagonal elements of any other output power spectral density matrix from any other filter bank in $\mathcal{U}_{\infty,M}$. That in turn means that the corresponding output sub-band variance vector corresponding to the diagonal power spectral density matrix majorizes the sub-band variance vector from any other output power spectral density. A filter bank that diagonalizes the input power spectral density matrix at each frequency is a coding gain optimal filter bank by Theorem 3.3.3.

Although we have discussed optimality of filter banks over the whole universe $\mathcal{U}_{\infty,M}$, majorization theory and Remark 4.3.1 allow us to discuss principal component optimality and coding gain optimality over subsets of $\mathcal{U}_{\infty,M}$. Consider a class of filter banks $\mathcal{C} \subseteq \mathcal{U}_{\infty,M}$. Denote by $\mathcal{S} \subset \mathbb{R}^M$ the *search space* [9] of output sub-band variance vectors from all the filter banks in $\mathcal{C}$. Suppose there is a vector $\boldsymbol{\sigma}_* \in \mathcal{S}$ that majorizes any other sub-band

variance vectors in S. Then by Remark 4.3.1 the filter bank with polyphase matrix denoted by $\boldsymbol{E}_*(\cdot)$ that outputs $\boldsymbol{\sigma}_*$ is a coding gain optimal filter bank over $\mathcal{C}$. What can we infer about the the compaction gain from filters in $\mathcal{C}$? Since $\boldsymbol{\sigma}_*$ majorizes any other variance vector in S, the compaction gain from $\boldsymbol{E}_*(\cdot)$ is also maximum over all filter banks in $\mathcal{C}$ – and to prove this assertion we just write the majorization relation for the first component.

In conclusion, for any class of filter banks $\mathcal{C} \subseteqq \mathcal{U}_{\infty,M}$ the majorization theory and the theory of optimal orthonormal filter banks guarantee that if a sub-band variance vector $\boldsymbol{\sigma}_*$ is a maximal element with respect to the majorization relation over the search space corresponding to the filter banks in $\mathcal{C}$ then $\boldsymbol{\sigma}_*$ is the output from an optimal coding gain (and optimal compaction gain) filter bank, $\boldsymbol{E}_*(\cdot)$. Moreover, the reconstruction error defined in Equation 4.1 is minimized when only the first P sub-bands are used to reconstruct the signal, i.e., $\boldsymbol{E}_*(\cdot)$ corresponds to a filter bank that is optimal in the principal component sense.

A rigorous definition of principal component filter banks can be formulated as:

Definition 4.3.3. [9] Consider a class of orthonormal uniform maximally down-sampled M-channel filter banks $\mathcal{C} \subseteqq \mathcal{U}_{\infty,M}$. A filter bank is a principal component filter bank for the class $\mathcal{C}$ if its sub-band variance vector majorizes the sub-band variance vector of any other filter in the class.

Remark 4.3.3.   • The majorization relation is a partial order because it is

110

reflexive, antisymmetric and transitive.

- The sub-band variance vector corresponding to a principal component filter bank with the sub-band variances ordered decreasingly is unique in $\mathcal{U}_{\infty,M}$. That is because the majorization relation defined on the set of sub-band variance vectors is antisymmetric.

- The output sub-band variance vectors from filters in $\mathcal{U}_{\infty,M}$ lie in a hyperplane of $\mathbb{R}_+^M$, for each input. That is because the sum of the output variances is constant by Proposition 3.2.1.

- Denote by $\mathcal{S}$ the set of all output sub-band variance vectors from a class $\mathcal{C} \subset \mathcal{U}_{\infty,M}$ and call it the *search space*

$$\mathcal{S} = \left\{ \boldsymbol{\sigma} \in \mathbb{R}_+^M : \boldsymbol{\sigma} = \mathrm{diag}\left(\int_0^{2\pi} \boldsymbol{S_{vv}}(e^{i\omega}) d\omega/(2\pi) \right), \ \ \boldsymbol{E}(\cdot) \in \mathcal{C} \right\}.$$

On $\mathcal{S}$ the majorization is a relation of total order if M is 1 or 2 and a partial order if M is greater than 2.

4.4 Construction of Infinite Impulse Response Principal Component Filter Banks

The approach to constructing optimal compaction filters in Theorem 3.4.1 can also be used to construct principal component filter banks. A valid compaction filter satisfies the Nyquist(M) condition. An optimum compaction filter maximizes its output variance (see Section 3.3). Therefore, in an optimum compaction orthonormal filter bank, there is one filter whose output

variance is maximum for the given input power spectral density, for all orthonormal filter banks in the class considered. For now the class of filters considered is $\mathcal{U}_{\infty,M}$.

From the definition of principal component filter banks it follows that a principal component filter bank in $\mathcal{U}_{\infty,M}$ contains an optimum compaction filter. In other words, a principal component filter bank is also an optimum compaction filter bank. This result is proved by writing the majorization relation for the first component of the output sub-band variance vector.

Optimizing coding gain is equivalent to minimizing the product of sub-band variances; the optimum compaction gain is achieved for a filter bank whose first component is the largest among all the first components of all the output variance vectors in the class considered. When sequential optimum compaction filters are constructed following the "peel-off" procedure in Theorem 3.4.1 the resulting ideal filter bank is coding gain optimal in $\mathcal{U}_{\infty,M}$. It is shown that the output sub-band variance of such a filter bank majorizes all the output variances from filter banks in $\mathcal{U}_{\infty,M}$. See Vaidyanathan [118].

Proposition 4.4.1. Vaidyanathan [118]. *Iterative construction of optimum compaction filters using Theorem 3.4.1 produces a principal component filter bank for $\mathcal{U}_{\infty,M}$.*

112

4.5 Optimality of Principal Component Filter Banks

Optimality of principal component filter banks is discussed in the presence of additive quantizer noise.

When a signal is approximated by its first P sub-bands, the reconstruction error is, as in Equation (4.1)

$$\epsilon(P) = \frac{1}{M} \sum_{k=P}^{M-1} \sigma_{v_k}^2, \qquad P \leq M.$$

This error may be generically expressed in terms of a continuous function g of the output sub-band variances [9, 115]:

$$g : \mathbb{R}_+ \setminus \{0\} \to \mathbb{R},$$

$$\phi(\boldsymbol{\sigma}) = \sum_{k=0}^{M-1} g(\sigma_{v_k}^2), \qquad \boldsymbol{\sigma} = [\sigma_{v_0}^2, \sigma_{v_1}^2, \ldots, \sigma_{v_{M-1}}^2]^T.$$

Optimality of a filter bank may be considered with respect to ϕ. If the objective function is Schur-convex (Schur-concave) and a maximum(minimum) of ϕ is sought then a principal component filter banks is optimal with respect to the criterion expressed by ϕ.

Consider a sub-band coder as in Figure 1.9 with M channels operated by an orthonormal filter bank. The input variance is σ_x^2 and the output sub-band variances are $\sigma_{v_k}^2$ and the average of the sub-band variances is equal to the input variance (Proposition 3.2.1). If the quantizers operate at high bit rate, the sub-band quantizer noise is given by the standard noise model:

$$\sigma_{q_k}^2 = c\, 2^{-b_k} \sigma_{v_k}^2, \qquad k = 0, 1, \ldots, M-1$$

and the given average bit rate is $b = \sum_{k=0}^{M-1} b_k$. Equation 1.14 gives the coding gain

$$G_{SBC} = \frac{(1/M) \sum_{k=0}^{M-1} \sigma_{v_k}^2}{(\prod_{k=0}^{M-1} \sigma_{v_k}^2)^{1/M}}$$

and it is maximum when the filter bank is a principal component filter bank – as shown in Section 4.3.

However, not all the quantization processes operate at high bit rate, hence it is desirable to discuss the optimality of principal component filter banks at arbitrary bit rates. For transform coders, the Karhunen-Loève transform – which is the principal component filter banks for the class of memoryless filter banks – is optimal even when the quantizers do not operate at high bit rates – as shown by Huang and Schultheiss [51] (see Section 2.5). Kirac and Vaidyanathan [61] propose a more general quantizer model and show that principal component filter banks in $\mathcal{U}_{\infty,M}$ are optimal in the presence of quantizers that operate at arbitrary bit rates. Optimality is understood in the coding gain sense.

With the more general model, the sub-band noise variances are

$$\sigma_{q_k}^2 = f(b_k)\sigma_{v_k}^2, \qquad k = 0, 1, \ldots, M-1. \tag{4.6}$$

There is a single quantization function for all the channels. This simplification is in general not true unless the input is Gaussian. The functions that model the quantizer should be independent of the filter bank; however, the quantizers are optimized to their input probability density function, which is in turn

influenced by the filter bank. When the input is Gaussian the output sub-band signals are also Gaussian and the quantizers and the filters can be chosen independently. Kirac and Vaidyanathan prove:

Theorem 4.5.1. *With a general quantizer model described by Equation (4.6), a principal component filter bank, if it exists, minimizes the reconstruction error of an orthonormal sub-band coder for all bit budgets and bit allocations.*

Remark 4.5.1. The optimization of the filter bank and the bit allocation are now decoupled. Even if the bit allocation strategy is not optimal, a principal component filter bank is optimal – with respect to the coding gain and mean-square reconstruction error.

Remark 4.5.2. Consider $f(b_k) = c$ constant over all channels. This is the equal bit allocation strategy. In this case the reconstruction error is, by Equation (4.1), $\epsilon = c\sigma_x^2$, independent of the choice of the filter bank. Hence orthonormal filter banks do not yield any coding improvement if equal bit allocation is employed.

From the previous theorem it follows that principal component filter banks over any class $\mathcal{C} \subset \mathcal{U}_{\infty,M}$ are optimal independent of the bit rates and the bit allocation strategy as long as the quantizer model relies on only one underlying quantization function.

4.6 New Result on Existence of Optimal Filter Banks in Finite Impulse Response Classes

The important results presented in the previous sections ensure the existence and provide the design of principal component filter banks in the unrestricted class $\mathcal{U}_{\infty,M}$ with infinite impulse response filters. The definition of principal component filter banks is applicable to any class of filters $\mathcal{C} \subseteqq \mathcal{U}_{\infty,M}$; however, there are no results on existence and design of finite impulse response principal component filter banks with more than two channels, except for memoryless filter banks for which the Karhunen-Loève transform is the principal component filter bank.

The coding gain is a universally used metric for the performance of a filter bank in a sub-band coding scheme. It is therefore reasonable to seek first a coding gain optimal finite impulse response filter bank in a class $\mathcal{U}_{\mu,\ M}$, with $\mu > 0$, before we investigate if such a coding gain optimal filter bank is also a principal component filter bank. We know that the converse is true: a principal component filter bank in a class $\mathcal{C} \subseteqq \mathcal{U}_{\infty,M}$ is a coding gain optimal filter bank over that class. We now show that in each finite impulse response class there exists a filter bank that renders the optimum coding gain over all the filters in the class. We give an existence proof and in Section 4.7.4 we outline an algorithm to find these optimal filters for both the real and the complex case.

In [119] Vaidyanathan and Akkarakaran review the theory and applications of orthonormal filter banks, including transform coders. In regard to

the class of orthonormal finite impulse response filter banks with more than two channels, they state (page 284):

" For this class there is no procedure to find the globally optimal finite impulse response orthonormal filter bank to maximize the coding gain, even under high bit-rate assumptions".

The next theorem provides the theoretical support for a successful algorithm to find coding gain optimal filter banks. It guarantees the existence of coding gain optimal finite impulse response orthonormal filter banks.

Theorem 4.6.1. *Let an input signal be such that its power spectral density is integrable and it vanishes at most on a set of measure zero. For any such signal there exists a filter bank that maximizes the coding gain over the class of causal finite impulse response orthonormal filter banks with a given finite McMillan degree and a fixed number of channels.*

Let the number of channels be a positive integer M. Consider the set $\mathcal{U}_{\mu,M}$ of causal finite impulse response orthonormal M-channel filter banks with McMillan degree μ. The corresponding polyphase matrices are unitary matrices when the argument z is on the unit circle.

Recall from Section 3.2.1 that the polyphase matrix of a perfect reconstruction uniform orthonormal M-channel filter bank with McMillan degree μ is:

$$\boldsymbol{E}(z) = \prod_{k=1}^{M} \left(\boldsymbol{I} + (-1 + z^{-1}) \boldsymbol{v}_k \boldsymbol{v}_k^{\dagger} \right) \boldsymbol{U}, \qquad \text{for } |z| = 1$$

117

where $\|\boldsymbol{v}_k\| = 1, \ k = 1, 2, \ldots, M$ and $\boldsymbol{U}$ is a unitary matrix. Filter banks are completely determined by their values on the unit circle $z = e^{i\omega}$ with $\omega \in [0, 2\pi]$; however, from now on in this section we only display ω instead of $e^{i\omega}$ in the representation of the polyphase matrix.

We prove the theorem with the aid of a few preliminary results.

Definition 4.6.1. [96]. Two norms N_1 and N_2 on a vector space X are said to be *equivalent* if there exists positive constants α and β such that

$$\alpha N_1 \leq N_2 \leq \beta N_1 .$$

Proposition 4.6.2. [25, 47]. *All norms on a finite dimensional vector space are equivalent.*

Theorem 4.6.3. [25, 47]. *In a finite-dimensional normed linear space, each closed and bounded set is compact.*

We now define the entities used in the proof of Theorem 4.6.1. We follow in general the notation from Golub and van Loan [43] and Horn and Johnson [47].

- The norm used for vectors is the Euclidean norm, unless otherwise specified.

- The subordinate 2-norm – also called the operator 2-norm, or simply the 2-norm [43, 47] is:
$$\|\boldsymbol{A}\|_2 = \sup_{\|\boldsymbol{x}\|=1} \|\boldsymbol{A}\boldsymbol{x}\| .$$

118

If a matrix norm is used without an index, it is understood that it is the subordinate 2-norm.

- The complex algebra of $M \times M$ complex matrices equipped with the 2-norm is denoted by $\mathbb{C}^{M \times M}$. Note that this is a Banach algebra [96]. We note that all norms on $\mathbb{C}^{M \times M}$ are equivalent by Proposition 4.6.2

- Denote the set of complex unit norm vectors of dimension M by $\mathcal{S}^{M-1}$ and the group of $M \times M$ unitary matrices by $U(M)$:

$$
\begin{aligned}
\mathcal{S}^{M-1} &= \left\{ \boldsymbol{v} \in \mathbb{C}^M : \|v\| = 1 \right\}, \\
U(M) &= \left\{ \boldsymbol{U} \in \mathbb{C}^{M \times M} : \boldsymbol{U}^\dagger \boldsymbol{U} = \boldsymbol{U} \boldsymbol{U}^\dagger = \boldsymbol{I} \right\}.
\end{aligned}
$$

These sets are compact because they are closed and bounded in finite dimensional linear spaces $\mathbb{C}^M$ and $\mathbb{C}^{M \times M}$, respectively [47].

- Let $C([0, 2\pi], \mathbb{C}^{M \times M})$ denote the Banach space of continuous matrix-valued functions equipped with the sup norm:

$$
\|F\|_{sup} = \sup_{\omega \in [0, 2\pi]} \|F(\omega)\|_2.
$$

- The set $(\mathcal{S}^{M-1})^\mu \times U(M)$ is compact in $(\mathbb{C}^M)^\mu \times \mathbb{C}^{M \times M}$ [25].

We will show that the coding gain is a continuous function defined on the compact set $(\mathcal{S}^{M-1})^\mu \times U(M)$ – see Figure 4.1.

Lemma 4.6.4. *Let D_k be compact sets in finite-dimensional normed linear spaces X_k, $k = 1, 2$ and let F_k be continuous maps from D_k to a Banach*

119

$$(\mathcal{S}^{M-1})^\mu \times U(M)$$

$$\downarrow \quad E$$

$$C([0,2\pi], U(M))$$

$$\downarrow \quad ES_{xx}E^\dagger$$

$$L^1([0,2\pi], \mathbb{C}^{M \times M})$$

$$\downarrow \quad \mathcal{I}$$

$$\mathbb{R}_+^{M \times M}$$

$$\downarrow \quad \mathrm{diag}$$

$$(0,1)^M \subset \mathbb{R}^M$$

$$\downarrow \quad CG$$

$$[1,\infty) \subset \mathbb{R}$$
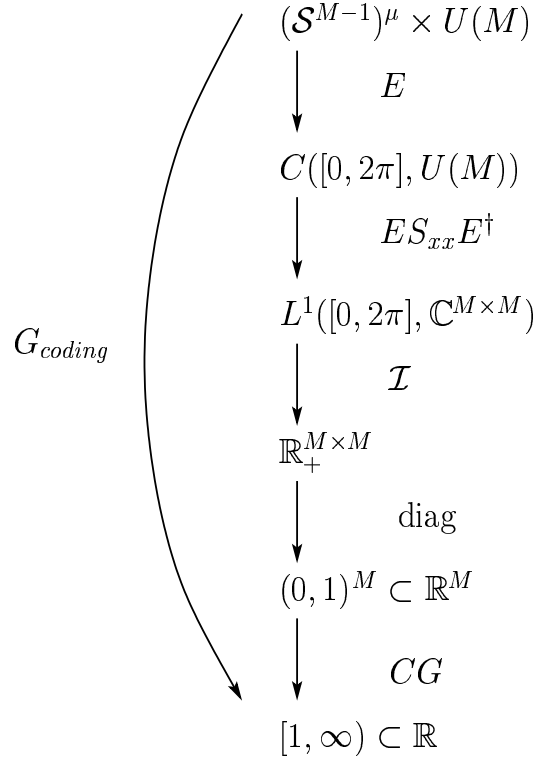
G_{coding}

Figure 4.1: Function diagram for coding gain

algebra Z. Then the tensor product

$$(F_1 \otimes F_2)(x_1, x_2) = F_1(x_1)F_2(x_2), \qquad x_k \in D_k, \;\; k = 1,2$$

is continuous.

Proof. Let a norm be defined on $X_1 \times X_2$ as:

$$\|(x_1, x_2)\|_\infty = \max\{\|x_1\|, \|x_2\|\}, \qquad \text{for all } (x_1, x_2) \in X_1 \times X_2$$

where the norms on the right hand side are on X_1 and X_2, respectively. Let

$$M = \max_{k=1,2}\{ \sup_{x_k \in D_k} \|F_k(x_k)\| \}.$$

Let $(x_1, x_2) \in D_1 \times D_2$ and $\epsilon > 0$. From the continuity of F_1 and F_2 there exists $\delta_k > 0$ such that for all $y_k \in D_k$ satisfying $\|x_k - y_k\| < \delta_k$, $k = 1, 2$

$$M \|F_k(x_k) - F_k(y_k)\| < \frac{\epsilon}{2} \,.$$

Let $\delta = \min\{\delta_1, \delta_2\}$. Then for all $\|(x_1, x_2) - (y_1, y_2)\|_\infty < \delta$ we have:

$$\|(F_1 \otimes F_2)(x_1, x_2) - (F_1 \otimes F_2)(y_1, y_2)\|$$
$$= \|F_1(x_1)F_2(x_2) - F_1(x_1)F_2(y_2) + F_1(x_1)F_2(y_2) - F_1(y_1)F_2(y_2)\|$$
$$\leq \|F_1(x_1)(F_2(x_2) - F_2(y_2))\| + \|(F_1(x_1) - F_1(y_1))F_2(y_2)\|$$
$$\leq \|F_1(x_1)\| \, \|F_2(x_2) - F_2(y_2)\| + \|F_2(x_2)\| \, \|F_1(x_1) - F_1(y_1)\|$$
$$< \frac{\epsilon}{2} + \frac{\epsilon}{2}$$
$$= \epsilon \,.$$

This shows that the tensor product $F_1 \otimes F_2$ is continuous. We used the result from [95] page 91: a continuous function defined from a compact set of a metric space to another metric space is uniformly continuous. $\qquad\square$

A similar result is stated in the next remark, with a proof that parallels the one from the previous lemma.

Remark 4.6.1. Let F_k be continuous maps from finite-dimensional normed linear spaces X_k, $k = 1, 2$ to a Banach algebra Z. Then the tensor product

$$F_1 \otimes F_2 :\to Z \,,$$

$$(F_1 \otimes F_2)(x_1, x_2) = F_1(x_1)F_2(x_2), \qquad x_k \in X_k, \ k = 1, 2$$

is continuous.

Remark 4.6.2. Note that by using the associativity of the multiplicative operation in Z, we infer the continuity of the tensor product of any finite number of functions that satisfy the hypothesis of the previous lemma.

Lemma 4.6.5. *Let $\boldsymbol{v} \in \mathcal{S}^{M-1}$. Define $V\boldsymbol{v} : [0, 2\pi] \to \mathbb{C}^{M \times M}$ by:*

$$V_{\boldsymbol{v}}(\omega) = \boldsymbol{I} + (-1 + e^{-i\omega})\boldsymbol{v}\boldsymbol{v}^{\dagger}.$$

Then $V_{\boldsymbol{v}}$ is an element of $C([0, 2\pi], \mathbb{C}^{M \times M})$.

Proof. Let $\epsilon > 0$ and $\alpha \in [0, 2\pi]$. Then there exists $\delta > 0$ such that for all $\omega \in [0, 2\pi]$ with $|\alpha - \omega| < \delta$ the inequality $|\sin \frac{\alpha - \omega}{2}| < \frac{\epsilon}{2}$ holds, by the continuity of the sine function on $[0, 2\pi]$. We now evaluate

$$
\begin{aligned}
\|V_{\boldsymbol{v}}(\alpha) - V_{\boldsymbol{v}}(\omega)\| &= \|(e^{-i\alpha} - e^{-i\omega})\boldsymbol{v}\boldsymbol{v}^{\dagger}\| \\
&\leq 2\left|\sin \frac{\alpha - \omega}{2}\right| \\
&< \epsilon.
\end{aligned}
$$

$\square$

Lemma 4.6.6. *The map $\boldsymbol{v} \mapsto V_{\boldsymbol{v}}$ is continuous from $\mathcal{S}^{M-1}$ to $C\left([0, 2\pi], \mathbb{C}^{M \times M}\right)$.*

Proof. Let $\boldsymbol{v}, \boldsymbol{w}$ in $\mathcal{S}^{M-1}$. Evaluate:

$$
\begin{aligned}
\|V_{\boldsymbol{v}} - V_{\boldsymbol{w}}\|_{sup} &= \sup_{\omega \in [0, 2\pi]} \|V_{\boldsymbol{v}}(\omega) - V_{\boldsymbol{w}}(\omega)\| \\
&= \sup_{\omega \in [0, 2\pi]} \|(-1 + e^{-i\omega})(\boldsymbol{v}\boldsymbol{v}^{\dagger} - \boldsymbol{w}\boldsymbol{w}^{\dagger})\| \\
&\leq 2\|\boldsymbol{v}(\boldsymbol{v}^{\dagger} - \boldsymbol{w}^{\dagger}) + (\boldsymbol{v} - \boldsymbol{w})\boldsymbol{w}^{\dagger}\| \\
&\leq 4\|\boldsymbol{v} - \boldsymbol{w}\|.
\end{aligned}
$$

We conclude that $\|V_{\boldsymbol{v}} - V_{\boldsymbol{w}}\|_{sup} \to 0$ as $\|\boldsymbol{v} - \boldsymbol{w}\| \to 0$. $\qquad\qquad\qquad\square$

Lemma 4.6.7. *Let $\boldsymbol{\theta} = (\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_\mu, \boldsymbol{U})$ be an element of $(\mathcal{S}^{M-1})^\mu \times U(M)$. The mapping*

$$\boldsymbol{\theta} \mapsto \boldsymbol{E}_{\boldsymbol{\theta}} = V_{\boldsymbol{v}_1} V_{\boldsymbol{v}_2} \cdots V_{\boldsymbol{v}_\mu} \boldsymbol{U}$$

is continuous from $(\mathcal{S}^{M-1})^\mu \times U(M)$ to $C([0, 2\pi], U(M))$, where $U(M) \subset \mathbb{C}^{M \times M}$.

Proof. Observe that the map $\boldsymbol{E}$ is the tensor product of the maps $V_{\boldsymbol{v}_k}$, $k = 1, 2, \ldots, \mu$ and $\boldsymbol{U}$, all defined on compact sets in finite-dimensional linear spaces and with values in the Banach algebra of $M \times M$ complex matrices. The result of the lemma is therefore a consequence of Lemma 4.6.4 and Remark 4.6.2. $\qquad\qquad\qquad\square$

Lemma 4.6.8. *Let $\boldsymbol{S}$ be in $L^1([0, 2\pi], \mathbb{C}^{M \times M})$. The mapping*

$$\boldsymbol{E} \mapsto \boldsymbol{E}\boldsymbol{S}\boldsymbol{E}^\dagger$$

is continuous from $C([0, 2\pi], U(M))$ to $L^1([0, 2\pi], \mathbb{C}^{M \times M})$.

Proof. Denote by s the value of the L^1 norm of $\boldsymbol{S}$ using the operator (subordinate) 2-norm for the matrices that are the values of $\boldsymbol{S}$ evaluated at each ω:

$$s = \int_0^{2\pi} \|\boldsymbol{S}(\omega)\| d\omega < \infty.$$

Let $\boldsymbol{E}, \boldsymbol{F}$ be two elements of $C([0, 2\pi], U(M))$ such that

$$s \|\boldsymbol{E} - \boldsymbol{F}\|_{sup} < \frac{\epsilon}{2}.$$

That means in turn:

$$s\left(\sup_{\omega\in[0,2\pi]}\|(\boldsymbol{E}-\boldsymbol{F})(\omega)\|\right) < \frac{\epsilon}{2}.$$

The same bound holds for the conjugate transpose.

Evaluate now the L^1 norm:

$$\int_0^{2\pi}\|\boldsymbol{E}(\omega)\boldsymbol{S}(\omega)\boldsymbol{E}^\dagger(\omega) - \boldsymbol{F}(\omega)\boldsymbol{S}(\omega)\boldsymbol{F}^\dagger(\omega)\|d\omega$$

$$= \int_0^{2\pi}\|\boldsymbol{E}(\omega)\boldsymbol{S}(\omega)(\boldsymbol{E}^\dagger(\omega)-\boldsymbol{F}^\dagger(\omega))$$

$$+(\boldsymbol{E}(\omega)-\boldsymbol{F}(\omega))\boldsymbol{S}(\omega)\boldsymbol{F}^\dagger(\omega)\|d\omega$$

$$\leq \int_0^{2\pi}\|\boldsymbol{E}(\omega)\boldsymbol{S}(\omega)(\boldsymbol{E}^\dagger(\omega)-\boldsymbol{F}^\dagger(\omega))\|d\omega$$

$$+\int_0^{2\pi}\|(\boldsymbol{E}(\omega)-\boldsymbol{F}(\omega))\boldsymbol{S}(\omega)\boldsymbol{F}^\dagger(\omega)\|d\omega$$

$$\leq \int_0^{2\pi}\|\boldsymbol{E}(\omega)\|\,\|\boldsymbol{S}(\omega)\|\,\|\boldsymbol{E}^\dagger(\omega)-\boldsymbol{F}^\dagger(\omega)\|d\omega$$

$$+\int_0^{2\pi}\|\boldsymbol{E}(\omega)-\boldsymbol{F}(\omega)\|\,\|\boldsymbol{S}(\omega)\|\,\|\boldsymbol{F}^\dagger(\omega)\|d\omega$$

$$\leq \left(\sup_{\omega\in[0,2\pi]}\|\boldsymbol{E}^\dagger(\omega)-\boldsymbol{F}^\dagger(\omega)\|\right)\int_0^{2\pi}\|\boldsymbol{S}(\omega)\|d\omega$$

$$+\left(\sup_{\omega\in[0,2\pi]}\|\boldsymbol{E}(\omega)-\boldsymbol{F}(\omega)\|\right)\int_0^{2\pi}\|\boldsymbol{S}(\omega)\|d\omega$$

$$< \frac{\epsilon}{2}+\frac{\epsilon}{2}$$

$$= \epsilon.$$

$\square$

Lemma 4.6.9. *If $\boldsymbol{E}$ is an element of $L^1([0,2\pi],\ \mathbb{C}^{M\times M})$ then the map*

$$\boldsymbol{E}\mapsto \mathfrak{I}(\boldsymbol{E}) = \int_0^{2\pi}\boldsymbol{E}(\omega)d\omega \quad \in\ \mathbb{C}^{M\times M}$$

124

is continuous.

Proof. Let $\epsilon > 0$ and let $\boldsymbol{E}, \boldsymbol{F} \in L^1([0, 2\pi], \mathbb{C}^{M \times M})$ be such that $\|\boldsymbol{E} - \boldsymbol{F}\|_{L^1} < \epsilon$, that is:

$$\int_0^{2\pi} \|\boldsymbol{E}(\omega) - \boldsymbol{F}(\omega)\| d\omega < \epsilon.$$

Evaluate:

$$\|\Im(\boldsymbol{E}) - \Im(\boldsymbol{F})\|$$
$$= \|\int_0^{2\pi} (\boldsymbol{E}(\omega) - \boldsymbol{F}(\omega)) \, d\omega\|$$
$$\leq \int_0^{2\pi} \|\boldsymbol{E}(\omega) - \boldsymbol{F}(\omega)\| d\omega$$
$$< \epsilon.$$

$\square$

Lemma 4.6.10. *If* $\boldsymbol{\Sigma} \in \mathbb{C}^{M \times M}$ *then the map*

$$\boldsymbol{\Sigma} \mapsto \boldsymbol{\sigma} = diag(\boldsymbol{\Sigma}) \in \mathbb{C}^M$$

is continuous.

Proof. The map diag is linear. Every linear map on a finite-dimensional normed linear space is continuous [25, 95]. $\square$

Lemma 4.6.11. *Let* $\boldsymbol{\sigma} = [\sigma_1, \sigma_2, \ldots, \sigma_M]^T \in (0, 1)^M \subset \mathbb{R}^M$. *The map*

$$\boldsymbol{\sigma} \mapsto CG(\boldsymbol{\sigma}) = \frac{1}{(\prod_{k=1}^M \sigma_k)^{1/M}}$$

is continuous.

Proof. The product of the components of the vector $\boldsymbol{\sigma}$ is the tensor product of uniformly continuous maps

$$F_k \quad : \quad \mathbb{R}^M \to \mathbb{R}\,,$$

$$F_k(\boldsymbol{\sigma}) \quad = \quad F_k([\sigma_1, \sigma_1, \dots, \sigma_M]^T) = \sigma_k\,,$$

restricted to $(0,1)^M \subset \mathbb{R}^M$. This tensor product is continuous – see Remark 4.6.2 and Remark 4.6.1. The map CG is a composition of three maps: a product of the components of the vector $\boldsymbol{\sigma}$, a root of order M and an inverse, all continuous. $\qquad\square$

We can now prove Theorem 4.6.1:

Assume that the input variance is normalized, that is, $\sigma_x^2 = 1$. Since the input power spectral density does not vanish on any set of positive measure, it follows that the output sub-band variances are positive, for all orthonormal filter banks in $\mathcal{U}_{\mu,M}$, and by Proposition 3.2.1 their average is 1:

$$0 < \sigma_{v_k}^2 < 1, \qquad k = 1, 2, \dots, M\,, \qquad\qquad (4.7)$$
$$\sum_{k=1}^{M} \sigma_{v_k}^2 = 1\,.$$

The input matrix $\boldsymbol{S_{xx}}$ is a power spectral density matrix. The coding gain is the ratio of the arithmetic and geometric mean of the output variances, given in Equation 3.13, hence it is greater than or equal to 1.

The coding gain is a real-valued continuous function defined on the compact set $(\mathcal{S}^{M-1})^\mu \times U(M)$. By the extreme value theorem (Cheney [25]

126

page 16), there exists a point in $\boldsymbol{\theta}_0 \in (\mathcal{S}^{M-1})^\mu \times U(M)$ such that

$$G_{coding}(\boldsymbol{\theta}_0) = \max_{\boldsymbol{\theta} \in (\mathcal{S}^{M-1})^\mu \times U(M)} G_{coding}(\boldsymbol{\theta}).$$

The functional diagram is shown in Figure 4.1.

The theorem assures existence of optimal filter banks only within the class of causal orthonormal filter banks with McMillan degree μ, not on the whole universe $\mathcal{U}_{\infty,M}$.

4.7 Study of Optimal Filter Banks for Autoregressive Input Signals

Motivation.

The question of existence of principal component filter banks for finite impulse response classes is of great interest among the digital signal processing communities. Most results in the literature follow [113] and for more than two channels the resulting filters are ideal unrealizable filters. An intense theoretical effort and design strategies were aimed at developing finite impulse response principal component filter banks. The authors of [62] claim that principal component filter banks do not exist for the class of finite impulse response three channel para-unitary filter banks with McMillan degree 1. Our study is motivated by their claim and since there is no indication in their paper on the procedure used to create the optimal filter banks with respect to the coding and compaction gains, we decided to reproduce their experiments [90]. It was this study that led to the Theorem 4.6.1 on existence of finite impulse

127

response optimal filter banks. We present the study and then comment on the results from [62].

4.7.1 The input signal

Consider an autoregressive AR(1) discrete time stochastic process $x : \mathbb{Z} \to \mathbb{C}$ described by:

$$x[n] = \rho x[n-1] + w[n], \qquad n \in \mathbb{Z} \tag{4.8}$$

where $w : \mathbb{Z} \to \mathbb{C}$ is a Gaussian white noise process with variance σ_w^2 and $|\rho| < 1$.

Section 1.2.2 gives all the details on the model of an autoregressive AR(1) signal and the autocorrelation and power spectral density matrices for the blocked version of this signal. Assume that the input variance is unity.

The blocked signal is input to an orthonormal causal filter bank with McMillan degree 1. The filter banks have two and three channels.

The corresponding analysis polyphase matrices are para-unitary (Equation (3.6)). The analysis polyphase matrix of a finite impulse response filter bank with McMillan degree one can be written as in Equations (3.7) and (3.8)

$$\begin{aligned} \boldsymbol{E}(z) &= \boldsymbol{V}(z)\boldsymbol{U}, \\ \boldsymbol{V}(z) &= \boldsymbol{I} + (-1 + z^{-1})\boldsymbol{v}\boldsymbol{v}^T, \qquad \|\boldsymbol{v}\| = 1 \end{aligned}$$

and also as

$$\boldsymbol{E}(z) = \boldsymbol{E}_0 + z^{-1}\boldsymbol{E}_1 . \tag{4.9}$$

128

Since $\boldsymbol{E}$ is para-unitary the terms $\boldsymbol{E}_0$ and $\boldsymbol{E}_1$ obey:

$$\det \boldsymbol{E}_0 = 0, \quad \det \boldsymbol{E}_1 = 0, \quad \boldsymbol{E}_0 \boldsymbol{E}_1 = 0. \tag{4.10}$$

The study is divided in two parts: two-channel filter banks and three channel filter banks.

4.7.2   Two channel para-unitary filter banks

The input autoregressive $\mathrm{AR}(1)$ signal is passed through a real finite impulse response orthonormal two-channel maximally down-sampled filter bank with McMillan degree of 1. The blocked input signal is

$$\boldsymbol{x}[n] = [\ x[2n]\ \ x[2n-1]\ ]^T, \qquad n \in \mathbb{Z}. \tag{4.11}$$

The analysis polyphase matrix can be expressed in terms of unitary matrices and dyadic-based matrices – see Equations (3.7) and (3.8):

$$\boldsymbol{V}(z) = \boldsymbol{I} + (-1 + z^{-1})\boldsymbol{v}\boldsymbol{v}^T, \qquad |z| = 1$$

where $\boldsymbol{I}$ is the identity matrix in $\mathbb{R}^{2\times2}$ and $\boldsymbol{v} \in \mathbb{R}^2$ is a unit-norm vector. We can describe the whole finite impulse response class with McMillan degree 1, $\mathcal{U}_{1,2}$ by taking:

$$\boldsymbol{v}: \quad [0, 2\pi] \to \mathbb{R}, \tag{4.12}$$

$$\boldsymbol{U}: \quad [0, 2\pi] \to \mathbb{R}^{2\times2},$$

$$\boldsymbol{V}: \quad [0, 2\pi] \times [0, 2\pi] \to \mathbb{C}^{2\times2},$$

129

$$
\begin{aligned}
\boldsymbol{v}(\alpha) &= [\cos\alpha \ \sin\alpha]^T, \\
\boldsymbol{U}(\theta) &= \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}, \\
\boldsymbol{V}(\alpha, e^{i\omega}) &= \begin{bmatrix} \sin^2\alpha + e^{-i\omega}\cos^2\alpha & (-1 + e^{-i\omega})\cos\alpha\sin\alpha \\ (-1 + e^{-i\omega})\cos\alpha\sin\alpha & \cos^2\alpha + e^{-i\omega}\sin^2\alpha \end{bmatrix}.
\end{aligned}
$$

The analysis polyphase matrix can be chosen either as $\boldsymbol{E}(z) = \boldsymbol{V}(z)\boldsymbol{U}$ or $\boldsymbol{E}(z) = \boldsymbol{U}\boldsymbol{V}(z)$. (Here we emphasize only the memory argument, z). Since $\boldsymbol{V}$ is periodic in the first argument with period π, $\boldsymbol{E}$ is also periodic:

$$
\boldsymbol{E}(\alpha + \pi, \theta, e^{i\omega}) = \boldsymbol{E}(\alpha, \theta, e^{i\omega}).
$$

Thus the whole $\mathcal{U}_{1,2}$ class has its polyphase matrices described by three arguments:

$$
\boldsymbol{E} \ : \ [0, \pi] \times [0, 2\pi] \times [0, 2\pi] \rightarrow \mathbb{C}^{2\times 2}.
$$

Note. It suffices to consider only the orthogonal matrices with determinant 1; the same values of the variances are obtained when the orthogonal matrix has the determinant -1, with the minus sign attached to $\cos\theta$ (Murnaghan [83]). The autocorrelation matrix is:

$$
\boldsymbol{R_{xx}}(k) = \begin{cases} \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}, & k = 0 \\[2ex] \rho^{2k} \begin{bmatrix} 1 & \rho \\ \rho^{-1} & 1 \end{bmatrix}, & k > 0 \\[2ex] \rho^{-2k} \begin{bmatrix} 1 & \rho^{-1} \\ \rho & 1 \end{bmatrix}, & k < 0 \end{cases}
$$

130

The input power spectral density matrix is

$$
\begin{aligned}
\boldsymbol{S_{xx}}(z) &= \boldsymbol{R_{xx}}(0) + \sum_{k=1}^{\infty} \boldsymbol{R_{xx}}(k)z^{-k} + \sum_{k=1}^{\infty} \boldsymbol{R_{xx}}(-k)z^{k} \\
&= \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} + \frac{\rho^2 z^{-1}}{1 - \rho^2 z^{-1}} \begin{bmatrix} 1 & \rho \\ \rho^{-1} & 1 \end{bmatrix} + \frac{\rho^2 z}{1 - \rho^2 z} \begin{bmatrix} 1 & \rho^{-1} \\ \rho & 1 \end{bmatrix}
\end{aligned}
$$

and it easily verifies that the power spectral density matrix is pseudo-circulant. The Karhunen-Loève transform matrix for the autoregressive AR(1) input with correlation coefficient ρ is a discrete Fourier transform matrix because the autocorrelation matrix is circulant [127]:

$$
\boldsymbol{U}_{KLT} = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}
$$

and it satisfies:

$$
\boldsymbol{U}_{KLT}\boldsymbol{R_{xx}}(0)\boldsymbol{U}_{KLT}^{T} = \begin{bmatrix} 1 + \rho & 0 \\ 0 & 1 - \rho \end{bmatrix}.
$$

For a correlation coefficient ρ of 0.9 the eigenvalues of the auto-covariance matrix are $\lambda_1 = 1.9$ and $\lambda_2 = 0.1$ and the theoretical coding gain as defined in Equation (3.13) is $G_{codingKLT} = 2.294157$.

The sum of the output variances is constant, 2, therefore the analysis of the first input variance suffices to determine the behaviour of the coding gain. It is the coding gain that reflects the performance of a filter bank; we express it simply as:

$$
CG(\alpha, \theta) = \frac{1}{\sqrt{\sigma_{v_0}^2(\alpha, \theta)\sigma_{v_1}^2(\alpha, \theta)}} = \frac{1}{\sqrt{\sigma_{v_0}^2(\alpha, \theta)(2 - \sigma_{v_0}^2(\alpha, \theta))}} \tag{4.13}
$$

and its partial derivatives:

$$\frac{\partial CG(\alpha, \theta)}{\partial \alpha} = \frac{-1}{2}[\sigma_{v_0}^2(2 - \sigma_{v_0}^2)]^{-3/2}(\sigma_{v_0}^2 - 1)\frac{\partial \sigma_{v_0}^2(\alpha, \theta)}{\partial \alpha}, \qquad (4.14)$$
$$\frac{\partial CG(\alpha, \theta)}{\partial \theta} = \frac{-1}{2}[\sigma_{v_0}^2(2 - \sigma_{v_0}^2)]^{-3/2}(\sigma_{v_0}^2 - 1)\frac{\partial \sigma_{v_0}^2(\alpha, \theta)}{\partial \theta}.$$

Note that values of 1 for the output variances correspond to the trivial case of coding gain of 1, i.e., no gain (that is: the filters are just down-sampling operations) and a value of 2 for $\sigma_{v_0}^2$ implies a null $\sigma_{v_1}^2$ which would contradict the orthonormality and equivalently the perfect reconstruction property of the filter bank.

The maximum coding gain occurs when the first variance attains an extreme value, either maximum or minimum and trivially the compaction gain behaves as the first variance with respect to the arguments $(\alpha, \theta) \in [0, \pi] \times [0, 2\pi]$. The expressions for the output variances are determined analytically with Maple and so are their critical points and the attributes of the critical points (saddle, extrema). There are two possibilities of factoring the polyphase matrix with respect to the order of the dyadic structure and the unitary transform and we study each case separately.

4.7.2.1 Case $\boldsymbol{E}(z) = \boldsymbol{V}(z)\boldsymbol{U}$

A first evaluation for the analysis polyphase matrix is:

$\boldsymbol{E}(\alpha, \theta, e^{i\omega}) = \boldsymbol{V}(\alpha, e^{i\omega})\boldsymbol{U}(\theta)$, with $\boldsymbol{U}$ and $\boldsymbol{V}$ from Equation (4.12).

The variances are:

$$\sigma_{v_0}^2(\alpha,\theta) = 1 - 0.684\cos^2\alpha\cos\theta\sin\theta + 0.684\cos^4\alpha\cos\theta\sin\theta$$
$$+1.8\cos\theta\sin\theta - 0.342\cos\alpha\sin\alpha\cos^2\theta$$
$$+0.684\cos^3\alpha\sin\alpha\cos^2\theta - 0.342\cos^3\alpha\sin\alpha$$

and

$$\sigma_{v_1}^2(\alpha,\theta) = 1 + 0.342\cos\alpha\sin\alpha\cos^2\theta + 0.684\cos^2\alpha\cos\theta$$
$$-1.8\cos\theta\sin\theta\sin\theta + 0.342\cos^3\alpha\sin\alpha$$
$$-0.684\cos^4\alpha\cos\theta\sin\theta - 0.684\cos^3\alpha\sin\alpha\cos^2\theta \, .$$

The next four equations show the periodicity of the variances:

$$\sigma_{v_k}^2(\alpha + \pi, \theta) = \sigma_{v_k}^2(\alpha,\theta), \tag{4.15}$$
$$\sigma_{v_k}^2(\alpha, \theta + \pi) = \sigma_{v_k}^2(\alpha,\theta), \qquad k = 0, 1.$$

and we restrict the study of the coding gain on the interval $[0,\pi] \times [0,\pi] \subset \mathbb{R}^2$. Table 4.1 shows critical values of the first variance and the values of the coding gain and the "Comments" column records attributes of the coding gain.

Remark 4.7.1. The minimum values of the coding gain are not listed. That is because the extreme values of the coding gain in the table are calculated from the extreme values of the first variance. When the first variance attains a minimum value, the coding gain attains a maximum value because the second variance reaches a maximum. The minimum value of the coding gain is reached when $\sigma_{v_0}^2 = \sigma_{v_1}^2$.

133

α	θ	$\sigma^2_{v_0}$	Coding gain	Comment
$\pi/4$	$\pi/4$	1.72900	1.46089	saddle
$3\pi/4$	$3\pi/4$	0.27100	1.46089	saddle
1.83973	.76397	1.92216	2.58538	maximum
1.30185	2.37761	0.07783	2.58538	maximum
2.87265	.80681	1.92216	2.58538	maximum
.26893	2.33477	0.07783	2.58538	maximum

Table 4.1: Critical points for variances and corresponding coding gain values

The first variance and the coding gain are shown in Figures 4.2 and 4.3 respectively. When the matrix $\boldsymbol{U}$ is chosen to be the constant ma-
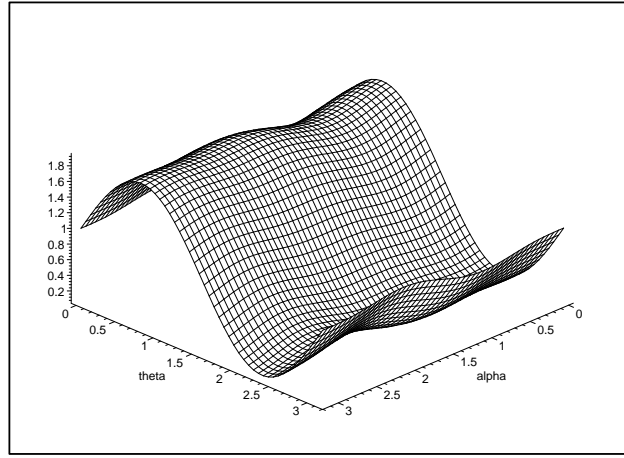
Figure 4.2: First variance : $\boldsymbol{E}(z) = \boldsymbol{V}(z)\boldsymbol{U}$

trix corresponding to the Karhunen-Loève transform of the input signal, i.e., $\boldsymbol{U}(\theta) = \boldsymbol{U}(\pi/4) = \boldsymbol{U}_{KLT}$ we obtain the maximum coding gain of 2.57284 at values of $\alpha \in \{-\arctan(11/4+\sqrt{3}+k\pi, k = 0, 1\}$ with corresponding variances of 1.92137 and .078624. When we compare these values with the values in Table 4.1, we conclude that the coding gain is not maximized when the matrix $\boldsymbol{U}$ is the Karhunen-Loève transform of the input and it can be concluded that
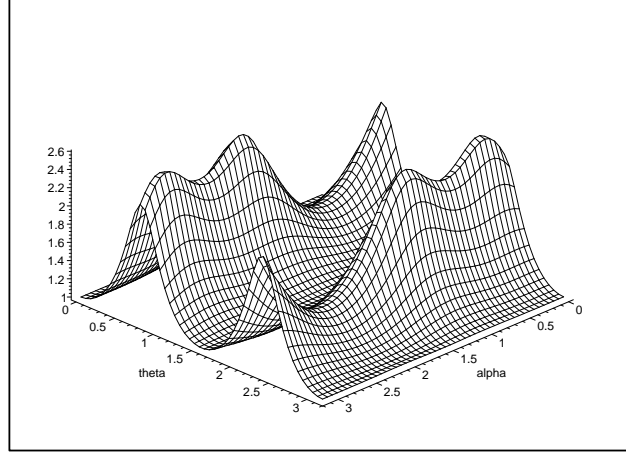
Figure 4.3: Coding gain: $\boldsymbol{E}(z) = \boldsymbol{V}(z)\boldsymbol{U}$

optimization by an iterative approach, one McMillan degree up at a time is not possible. Figures 4.4, 4.5, 4.6 and 4.7 show the first variance and the coding gain and their derivatives for this particular choice of $\boldsymbol{E}(z) = \boldsymbol{V}(z)\boldsymbol{U}_{KLT}$.
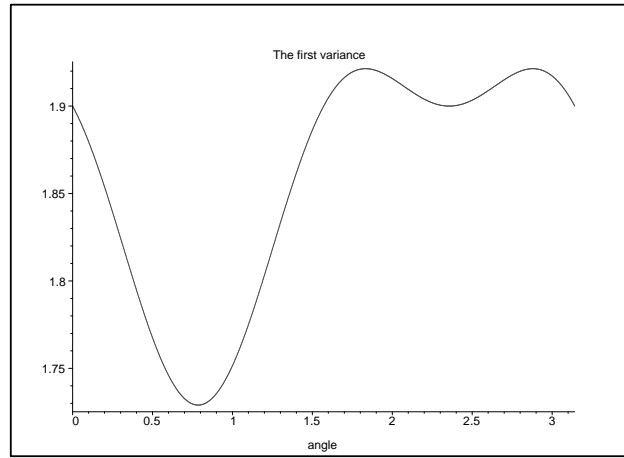

Figure 4.4: First variance: $\boldsymbol{E}(z) = \boldsymbol{V}(z)\boldsymbol{U}_{KLT}$

135
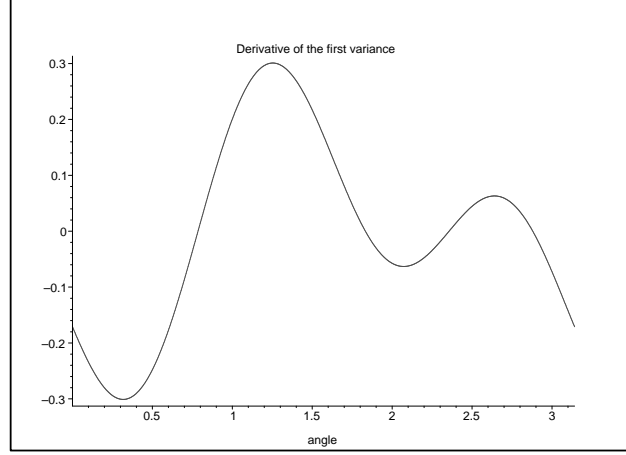
Figure 4.5: Derivative of the first variance: $\boldsymbol{E}(z) = \boldsymbol{V}(z)\boldsymbol{U}_{KLT}$

4.7.2.2   Case $\boldsymbol{E}(z) = \boldsymbol{U}\boldsymbol{V}(z)$

In terms of its factors, the analysis polyphase matrix is: $\boldsymbol{E}(\alpha, \theta, e^{i\omega}) = \boldsymbol{U}(\theta)\boldsymbol{V}(\alpha, e^{i\omega})$, with $\boldsymbol{U}$ and $\boldsymbol{V}$ as in Equation (4.12). The variances are:

$$
\begin{aligned}
\sigma_{v_0}^2(\alpha, \theta) &= 1 - 0.342\cos^3\alpha\sin\alpha + 0.342\cos\alpha\sin\alpha \\
&\quad + 1.458\cos\theta\sin\theta + 0.684\cos^3\alpha\sin\alpha\cos^2\theta \\
&\quad - 0.342\cos\alpha\sin\alpha\cos^2\theta + 1.026\cos^2\alpha\cos\theta\sin\theta \\
&\quad - 0.684\cos^4\alpha\cos\theta\sin\theta \,,
\end{aligned}
$$

and

$$
\begin{aligned}
\sigma_{v_1}^2(\alpha, \theta) &= 1 - 1.458\cos\theta\sin\theta - 0.684\cos^3\alpha\sin\alpha\cos^2\theta \\
&\quad + 0.684\cos^4\alpha\cos\theta\sin\theta - 1.026\cos^2\alpha\cos\theta\sin\theta \\
&\quad + 0.684\cos\alpha\sin\alpha\cos^2\theta - 0.342\cos\alpha\sin\alpha \\
&\quad + 0.342\cos^3\alpha\sin\alpha \,.
\end{aligned}
$$

136
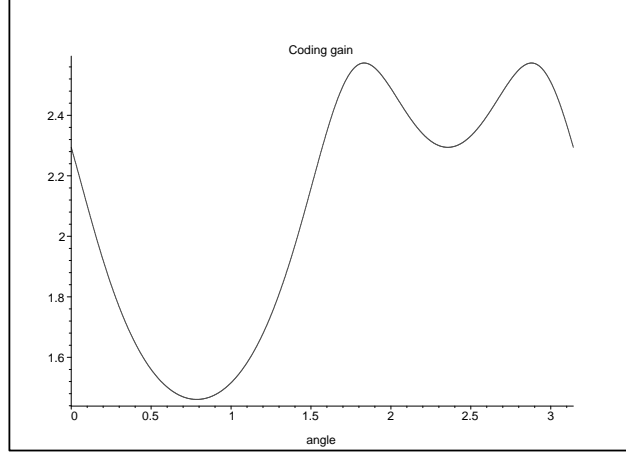
Figure 4.6: Coding gain: $\boldsymbol{E}(z) = \boldsymbol{V}(z)\boldsymbol{U}_{KLT}$
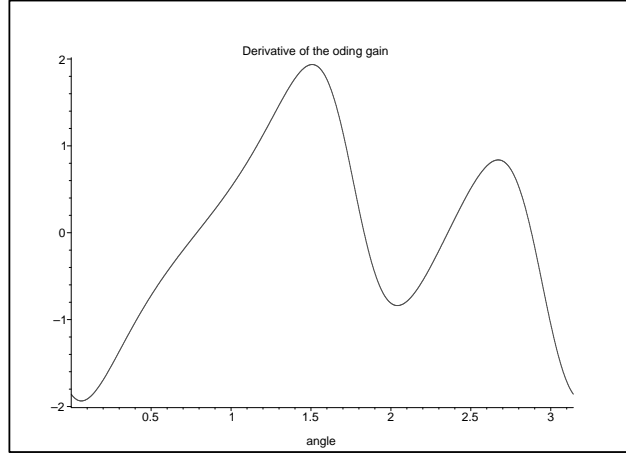
Figure 4.7: Derivative of the coding gain: $\boldsymbol{E}(z) = \boldsymbol{V}(z)\boldsymbol{U}_{KLT}$

The output variances satisfy Equations (4.15) and are periodic in both arguments with period π. We list in Table 4.2 the critical points $(\alpha, \theta) \in [0, \pi) \times [0, \pi)$. These critical points and their attributes (saddle and extrema) are determined analytically with Maple. The "Comments" column refers to attributes of the coding gain. Remark 4.7.1 justifies the absence of minimum

137

values for the coding gain in the table. The first variance and the coding gain

α	θ	$\sigma_{v_0}^2$	Coding gain	Comment
0	$\pi/4$	1.72900	1.46089	saddle
0	$3\pi/4$	0.10000	2.29415	saddle
$\pi/2$	$\pi/4$	1.729	1.46089	saddle
$\pi/2$	$3\pi/4$	0.27100	1.46089	saddle
.53787	.80681	1.92216	2.58538	maximum
2.60371	.76397	1.92216	2.58538	maximum
.53787	2.37761	0.07783	2.58538	maximum
2.60371	2.33477	0.07783	2.58538	maximum

Table 4.2: Critical points for variances and corresponding coding gain values

are shown in Figures 4.8 and 4.9 respectively. The maximum value of the cod-
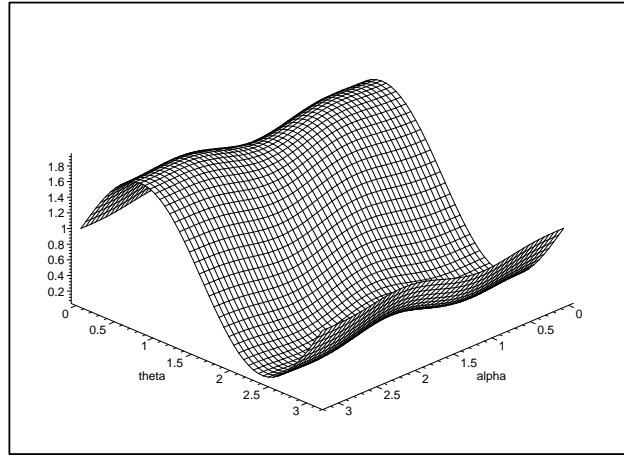
Figure 4.8: The first variance: $\boldsymbol{E}(z) = \boldsymbol{U}\boldsymbol{V}(z)$

ing gain is the same as for the previous case $\boldsymbol{E}(z) = \boldsymbol{V}(z)\boldsymbol{U}$ but it is attained at different points. We next show that at the points where the coding gain is maximized the matrix $\boldsymbol{U}(\theta)$ is the Karhunen-Loève transform of its input [118, 117, 80], i.e., we verify that $\boldsymbol{U}$ diagonalizes the auto-covariance matrix of the signal $\boldsymbol{w}$, denoted $\boldsymbol{R_{ww}}(0)$ (see Figure 4.10). In polyphase representation

138
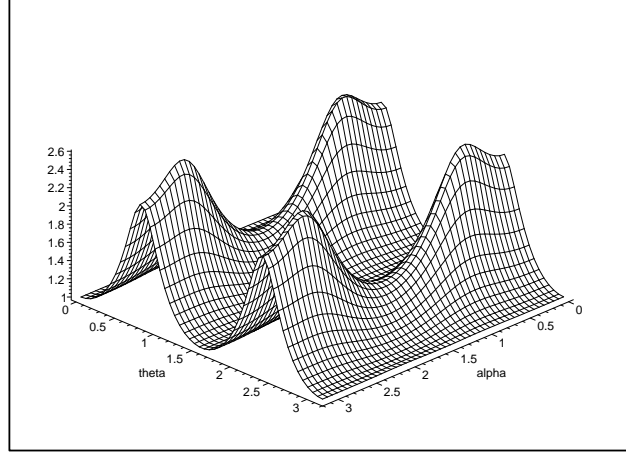
Figure 4.9: Coding gain: $\boldsymbol{E}(z) = \boldsymbol{U}\boldsymbol{V}(z)$

Analysis polyphase dyadic structure followed by unitary transform
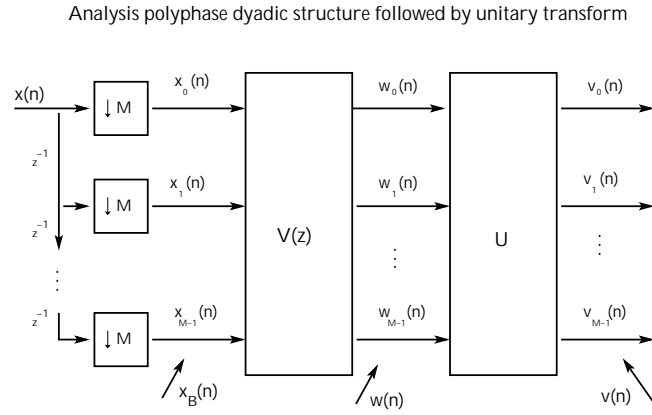
Figure 4.10: Factorization of the polyphase matrix, $\boldsymbol{E}(\text{z}) = \boldsymbol{U}\boldsymbol{V}(\text{z})$

the dyadic component $\boldsymbol{V}$ can be decomposed as:

$$\boldsymbol{V}(\alpha, z) = \boldsymbol{V}_0(\alpha) + z^{-1}\boldsymbol{V}_1(\alpha) \tag{4.16}$$

139

with

$$\boldsymbol{V}_0(\alpha) \;=\; \begin{bmatrix} \sin^2\alpha & -\cos\alpha\sin\alpha \\ -\cos\alpha\sin\alpha & \cos^2\alpha \end{bmatrix},$$

$$\boldsymbol{V}_1(\alpha) \;=\; \begin{bmatrix} \cos^2\alpha & \cos\alpha\sin\alpha \\ \cos\alpha\sin\alpha & \sin^2\alpha \end{bmatrix}.$$

Note that $\boldsymbol{V}_0$ and $\boldsymbol{V}_1$ are real and symmetric matrices. The output $\boldsymbol{w}$ is a function of the angle $\alpha \in [0,\pi)$ and in the polyphase representation $\boldsymbol{w}[n]$ is :

$$\boldsymbol{w}[n] = \boldsymbol{V}_0\boldsymbol{x}[n] + \boldsymbol{V}_1\boldsymbol{x}[n-1] \qquad (4.17)$$

and with $\boldsymbol{x}[n]$ from Equation (4.11); the auto-covariance is

$$
\begin{aligned}
\boldsymbol{R_{ww}}(0) \;&=\; E\{\boldsymbol{w}[n]\,\boldsymbol{w}^\dagger[n]\} \\[4pt]
&=\; E\{[\boldsymbol{V}_0\boldsymbol{x}[n] + \boldsymbol{V}_1\boldsymbol{x}[n-1]][\boldsymbol{V}_0\boldsymbol{x}[n] + \boldsymbol{V}_1\boldsymbol{x}[n-1]]^\dagger\} \\[4pt]
&=\; \boldsymbol{V}_0 E\{\boldsymbol{x}[n]\boldsymbol{x}^\dagger[n]\}\boldsymbol{V}_0^T + \boldsymbol{V}_0 E\{\boldsymbol{x}[n]\boldsymbol{x}^\dagger[n-1]\}\boldsymbol{V}_1^T \\[4pt]
&\quad +\boldsymbol{V}_1 E\{\boldsymbol{x}[n-1]\boldsymbol{x}^\dagger[n]\}\boldsymbol{V}_0^T + \boldsymbol{V}_1 E\{\boldsymbol{x}[n-1]\boldsymbol{x}^\dagger[n-1]\}\boldsymbol{V}_1^T \\[4pt]
&=\; \boldsymbol{V}_0 \boldsymbol{R_{xx}}(0)\boldsymbol{V}_0^T + \boldsymbol{V}_0 \boldsymbol{R_{xx}}(1)\boldsymbol{V}_1^T \\[4pt]
&\quad +\boldsymbol{V}_1 \boldsymbol{R_{xx}}(-1)\boldsymbol{V}_0^T + \boldsymbol{V}_1 \boldsymbol{R_{xx}}(0)\boldsymbol{V}_1^T .
\end{aligned}
$$

Recall that the autocorrelation matrices are:

$$\boldsymbol{R_{xx}}(0) \;=\; \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix},$$

$$\boldsymbol{R_{xx}}(1) \;=\; \rho^2 \begin{bmatrix} 1 & \rho \\ \rho^{-1} & 1 \end{bmatrix},$$

$$\boldsymbol{R_{xx}}(-1) \;=\; \begin{bmatrix} 1 & \rho^{-1} \\ \rho & 1 \end{bmatrix}.$$

140

The auto-covariance matrix of $\boldsymbol{w}$ evaluated at each argument α, $\boldsymbol{R_{ww}}(0)(\alpha)$ is:

$$\begin{bmatrix} -0.342\cos\alpha\sin^3\alpha + 1 & -0.342\cos^4\alpha + 0.513\cos^2\alpha + 0.729 \\ -0.342\cos^4\alpha + 0.513\cos^2\alpha + 0.729 & 0.342\cos\alpha\sin^3\alpha + 1 \end{bmatrix}.$$

We verified analytically with Maple that at all the critical points the matrix $\boldsymbol{U}$ diagonalizes the matrix $\boldsymbol{R_{ww}}(0)$, the auto-covariance of its input. The eigenvalues obtained are the extreme values for the variances rendering maximum coding gain or the variances corresponding to other critical points (saddle or minimum for coding gain). This behaviour of $\boldsymbol{U}$ acting upon $\boldsymbol{R_{ww}}(0)$ is expected because $\boldsymbol{U}$ is now a simple transform coder and the maximum coding gain is rendered by the Karhunen-Loève transform of the input to $\boldsymbol{U}$, i.e. when $\boldsymbol{U}$ diagonalizes $\boldsymbol{R_{xx}}(0)$. This is just a reassuring result; the search for the principal component filter bank is still an open problem since we have to design the dyadic-based structure $\boldsymbol{V}$ of the principal component filter bank polyphase matrix. Once $\boldsymbol{V}$ is known together with the input autocorrelation or power spectral density, finding $\boldsymbol{U}$ is an easy task. Numerically, we find $\boldsymbol{U}$ and the optimal filter bank guaranteed by Theorem 4.6.1 by:

- generate $\boldsymbol{V}$ as an expression of a vector $\boldsymbol{\alpha} \in [0,\pi]^\mu \subset \mathbb{R}^\mu$, with μ being the McMillan degree

- for each $\boldsymbol{\alpha}$ diagonalize the matrix $\boldsymbol{R_{ww}}(0)$ which depends on $\boldsymbol{\alpha}$. Select the largest eigenvalues – these are produced by the sought unitary matrix $\boldsymbol{U}$ that is a factor of the principal component filter bank polyphase matrix.

This procedure is an effective algorithm whose success is guaranteed by Theorem 4.6.1. When the power spectral density input is known as a function of frequency, Equations (3.11), (3.12) and (4.12) can be used to determine analytically the optimal filter bank – and in this case it is the principal component filter bank, too. There is only the practical limitation on the McMillan degree of the filter bank sought. The coding gain increases with the McMillan degree. For practical purposes, in applications where the input can be modeled as an autoregressive AR(1) signal, a ratio of the output variances of 1.922/0.078 as we obtain in our study is very good. Filter banks with order one polyphase render very good coding gain for this type of signals.

4.7.3 Three-channel filter banks

The AR(1) signal is input to a real finite impulse response orthonormal uniform three channel maximally down-sampled filter bank with McMillan degree 1. The blocked version of the input signal is:

$$\boldsymbol{x}[n] = [x[3n] \ \ x[3n-1] \ \ x[3n-2]]^T, \qquad n \in \mathbb{Z}.$$

We express the analysis polyphase matrix as in Section 3.2, Equations (3.7) and (3.8):

$$\boldsymbol{E}(z) \ = \ \boldsymbol{V}(z)\boldsymbol{U},$$
$$\boldsymbol{V}(z) \ = \ \boldsymbol{I} + (-1 + z^{-1})\boldsymbol{v}\boldsymbol{v}^T, \qquad |z| = 1.$$

The matrix $\boldsymbol{U}$ is unitary, $\boldsymbol{I}$ is the identity matrix in $\mathbb{R}^{3\times3}$ and $\boldsymbol{v} \in \mathbb{R}^3$ is a unit-norm vector. The whole finite impulse response class $\mathcal{U}_{1,3,O}$ can be described

142

by:

$$\boldsymbol{v} : [0, \pi] \times [0, 2\pi] \to \mathbb{R}^3, \tag{4.18}$$

$$\boldsymbol{U} : [0, 2\pi] \times [0, \pi] \times [0, 2\pi] \to \mathbb{R}^{3\times 3},$$

$$\boldsymbol{V} : [0, \pi] \times [0, 2\pi] \times [0, 2\pi] \to \mathbb{C}^{3\times 3},$$

$$\boldsymbol{v}(\alpha, \beta) = [\cos\alpha \quad \sin\alpha\cos\beta \quad \sin\alpha\sin\beta]^T,$$

$$\boldsymbol{V}(\alpha, \beta e^{i\omega}) = \boldsymbol{I} + (-1 + e^{-i\omega})\boldsymbol{v}(\alpha, \beta)\boldsymbol{v}^T(\alpha, \beta),$$

$$\boldsymbol{U}(\phi, \psi, \theta) = \begin{bmatrix} U_{11}(\phi, \psi, \theta) & U_{12}(\phi, \psi, \theta) & U_{13}(\phi, \psi, \theta) \\ U_{21}(\phi, \psi, \theta) & U_{22}(\phi, \psi, \theta) & U_{23}(\phi, \psi, \theta) \\ U_{31}(\phi, \psi, \theta) & U_{32}(\phi, \psi, \theta) & U_{33}(\phi, \psi, \theta) \end{bmatrix}$$

where

$$U_{11}(\phi, \psi, \theta) = \cos\psi\cos\phi - \cos\theta\sin\phi\sin\psi$$

$$U_{12}(\phi, \psi, \theta) = \cos\psi\sin\phi + \cos\theta\cos\phi\sin\psi$$

$$U_{13}(\phi, \psi, \theta) = \sin\psi\sin\theta$$

$$U_{21}(\phi, \psi, \theta) = -\sin\psi\cos\phi - \cos\theta\sin\phi\cos\psi$$

$$U_{22}(\phi, \psi, \theta) = -\sin\psi\sin\phi + \cos\theta\cos\phi\cos\psi$$

$$U_{23}(\phi, \psi, \theta) = \cos\psi\sin\theta$$

$$U_{31}(\phi, \psi, \theta) = -\sin\theta\sin\phi$$

$$U_{32}(\phi, \psi, \theta) = -\sin\theta\cos\phi$$

$$U_{33}(\phi, \psi, \theta) = \cos\theta.$$

Denote

$$D_E = [0, \pi] \times [0, 2\pi]^2 \times [0, 2\pi] \times [0, \pi] \times [0, 2\pi] \subset \mathbb{R}^6.$$

143

Thus the analysis polyphase matrix of filter banks in the class $\mathcal{U}_{1,3,O}$ is a function:

$$\boldsymbol{E} : D_E \to \mathbb{C}^{3\times 3}.$$

The autocorrelation matrix is a Toeplitz matrix:

$$\boldsymbol{R_{xx}}(k) = \begin{cases} \begin{bmatrix} 1 & \rho & \rho^2 \\ \rho & 1 & \rho \\ \rho^2 & \rho & 1 \end{bmatrix}, & k = 0 \\[4mm] \rho^{3k} \begin{bmatrix} 1 & \rho & \rho^2 \\ \rho^{-1} & 1 & \rho \\ \rho^{-2} & \rho^{-1} & 1 \end{bmatrix}, & k > 0 \\[4mm] \rho^{-3k} \begin{bmatrix} 1 & \rho^{-1} & \rho^{-2} \\ \rho & 1 & \rho^{-1} \\ \rho^2 & \rho & 1 \end{bmatrix}, & k < 0 \end{cases}.$$

The input power spectral density matrix is:

$$\boldsymbol{S_{xx}}(z) = \begin{bmatrix} 1 & \rho & \rho^2 \\ \rho & 1 & \rho \\ \rho^2 & \rho & 1 \end{bmatrix} + \frac{\rho^3 z^{-1}}{1 - \rho^3 z^{-1}} \begin{bmatrix} 1 & \rho & \rho^2 \\ \rho^{-1} & 1 & \rho \\ \rho^{-2} & \rho^{-1} & 1 \end{bmatrix}$$
$$+ \frac{\rho^3 z}{1 - \rho^3 z} \begin{bmatrix} 1 & \rho^{-1} & \rho^{-2} \\ \rho & 1 & \rho^{-1} \\ \rho^2 & \rho & 1 \end{bmatrix}.$$

and it can easily be verified that it is pseudo-circulant. Next, we determine the Karhunen-Loève transform matrix for the AR(1) signal with correlation coefficient 0.9. We expect to have a matrix different from a discrete Fourier transform matrix because the autocorrelation matrix is not circulant as it was in the previous case of two-channel filters. Before the Gramm-Schmidt

orthogonalization, the eigenvectors of the auto-covariance matrix are

$$\boldsymbol{u}_1 = [1, -\frac{9}{20} + \frac{1}{20}\sqrt{881}, 1]^T,$$

$$\boldsymbol{u}_2 = [1, -\frac{9}{20} - \frac{1}{20}\sqrt{881}, 1]^T,$$

$$\boldsymbol{u}_3 = [-1, 0, 1]^T.$$

with corresponding eigenvalues

$$\lambda_1 = \frac{281}{200} + \frac{9}{200}\sqrt{881} \approx 2.74067,$$

$$\lambda_2 = \frac{19}{100} = 0.19,$$

$$\lambda_3 = \frac{281}{200} - \frac{9}{200}\sqrt{881} \approx .06932.$$

The Karhunen-Loève transform is:

$$\boldsymbol{U}_{KLT} = \begin{bmatrix} .57079 & .59024 & .57079 \\ -.70710 & 0. & .70710 \\ .41736 & -.80722 & .41736 \end{bmatrix}$$

and it diagonalizes the auto-covariance matrix $\boldsymbol{R_{xx}}(0)$. We note that the Karhunen-Loève transform matrix is not a discrete Fourier transform matrix as it was the case for two channels and that is because the autocorrelation matrix is not circulant anymore – it is just Toeplitz. The coding gain corresponding to the Karhunen-Loève transform as defined in Equation (3.13) is: $G_{codingKLT} = 3.02573$. The output variances are

$$\sigma_{v_k}^2(\cdot) = \frac{1}{2\pi} \int_0^{2\pi} [\boldsymbol{E}(e^{i\omega}, \cdot)\boldsymbol{S_{xx}}(e^{i\omega})\boldsymbol{E}^\dagger(e^{i\omega}, \cdot)]_{kk} d\omega \qquad (4.19)$$

145

and they satisfy

$$\sigma^2_{v_k}(\alpha, \beta, \phi, \theta, \psi) = \sigma^2_{v_k}(\alpha, \beta + \pi, \phi, \theta, \psi), \tag{4.20}$$

$$\sigma^2_{v_k}(\alpha, \beta, \phi, \theta, \psi) = \sigma^2_{v_k}(\alpha + \pi, \beta, \phi, \theta, \psi + \pi), \tag{4.21}$$

$$k = 0, 1, 2.$$

The sum of the output variances is constant 3. The output variances and the coding gain are C^∞ trigonometric polynomials in five arguments. The maximum value of the coding gain does not depend on the order $\boldsymbol{E}(z, \cdot) = \boldsymbol{V}(z, \cdot)\boldsymbol{U}(\cdot)$ or $\boldsymbol{E}(z, \cdot) = \boldsymbol{U}(\cdot)\boldsymbol{V}(z, \cdot)$. We determine the output variances and the coding gain as follows:

1. use a symbolic computation package to integrate symbolically Equation (4.19); the output variances are now functions of the angles $\alpha, \beta, \theta, \phi$, and ψ

2. use numerical methods to calculate the variances.

The results are in the next table:

α	β	ϕ	θ	ψ	G_{cod}	G_{comp}	Comment
					3.0257	2.7406	KLT gain
0.3620	0.3324	6.3065	0.7745	2.4886	3.4355	2.7978	Max G_{cod}
1.2446	0.2971	0.5826	4.4324	2.5306	3.2907	2.7984	Max G_{comp}

Table 4.3: Coding and compaction gain values

There are other values of the arguments where the maximum compaction gain and the maximum coding gain are attained as shown by Equations (4.20) and here we list the first ones found.

146

The results show that there is a filter bank that renders a maximum coding gain but the corresponding compaction gain is not maximized. Theorem 3.3.3 on necessary and sufficient conditions for optimality prompts us to investigate if the output power spectral density matrix given by Equation (3.11) is diagonal. We check numerically and also analytically if the output power spectral density matrix is diagonal. There is no value of the argument $\omega \in [0, 2\pi]$ at which $\boldsymbol{S_{vv}}(e^{i\omega})$ is diagonal, therefore the filter we found to render the maximum coding gain is not a maximum compaction gain filter. However, this result does not contradict Theorem 4.6.1 on the existence of optimal finite impulse response filter banks , nor does it contradict the Definition 4.3.3 of principal component filter banks.

There are no principal component filter banks in the class of three-channel finite impulse response orthonormal filter banks with autoregressive AR(1) input signals.

This study is motivated by the results of Kirac and Vaidyanathan [62]. We correct their counterexample and support their conclusion. A careful reading of [62] reveals that the example is set improperly: In the study the three-dimensional unit norm vector used to construct the analysis polyphase matrix is $v = [\cos \alpha \ \sin \alpha \ 0]^T$. This choice of the vector reduces the study of a three -channel filter bank to a two-dimensional problem. Indeed when we use their vector for the construction of the generalized Householder matrix we observe that one component of the output variance vector is constant. Here, we use a proper representation of the unit norm vector (see Cheney and Light [26])

147

and determine analytically the output variance vector as a function of the arguments $\alpha, \beta, \phi, \theta, \psi$ listed in Equations (4.18). In the end, however, we reach the same conclusion as in [62]: there are no principal component filter banks for finite impulse response three channel para-unitary filter banks of degree 1 with AR(1) input.

4.7.4 Algorithm outline

The existence of coding gain optimal finite impulse response filter banks from $\mathcal{U}_{\mu,M}$ is guaranteed by Theorem 4.6.1. In practice the complexity of the design of such optimal filter banks depends on the input power spectral density and the size of the desired filter bank: number of channels and the McMillan degree. In Section 4.7.2 we found the coding gain optimal filter bank analytically. We now describe such an algorithm and then discuss its practicalities.

Suppose an input signal is given and the filters sought have M channels and McMillan degree μ. The coefficients of the filters are complex.

1. Determine the input $M \times M$ power spectral density matrix as a function of $z = e^{i\omega}$, with ω in $[0, 2\pi]$. This can be a challenge: simple expressions are desired but often the only option is the estimate of the periodogram.

2. Find a representation of the polyphase matrix $\boldsymbol{E}$ as a function of the argument $(z, \boldsymbol{\theta})$ where $z = e^{i\omega}$, $\omega \in [0, 2\pi]$ and $\boldsymbol{\theta}$ is tuple of polar coordinates that describe unit-norm vectors and unitary matrices.

148

3. Integrate Equation 3.12 with respect to $\omega \in [0, 2\pi]$ The output variances and the coding gain are functions of $\boldsymbol{\theta}$.

4. Determine the argument $\boldsymbol{\theta}_0$ for which the maximum value of the coding gain – or equivalently the minimum product of the output variances – is attained. The argument $\boldsymbol{\theta}_0$ describe the optimum filter bank.

In our study we carry all these steps analytically for two-channel filter banks. We use the symbolic computation package from Maple. For small size problems this method is recommended because it avoids the numerical integration in Step 3. For the three-channel filter banks it was possible to use the symbolic computation for Steps 1,2 and 3; we resort to numerical evaluation of the coding gain in Step 4. The expression of the power spectral density dictates if Step 3 can be computed symbolically.

We now discuss the number of polar coordinates used to parameterize the polyphase matrix. Real filter banks are designed for most practical applications. For the representation of real unit norm vectors v_k and the orthogonal matrix $\boldsymbol{U}$ the total number of parameters that describe the polyphase matrix $\boldsymbol{E}$ is

$$N_{real} = \mu(M - 1) + M(M - 1)/2\,. \tag{4.22}$$

For complex filters the number of polar coordinates is

$$N_{complex} = 2\mu(M - 1) + M^2 + 1\,. \tag{4.23}$$

The matrix $\boldsymbol{U}$ is complex unitary and a number of $M^2 + 1$ parameters are required for its representation.

149

Our study for $M = 2$ and $M = 3$ shows the parameterization for real unit norm vectors and orthogonal matrices. For the implementation with higher values of the number of channels M we refer to Cheney and Light [26], Murnaghan [83] and Vaidyanathan [117].

One may argue that the algorithm has a limited area of applications in the sense that it is suited for lower number of channels and/or low McMillan degree; we emphasize that the two-channel filter bank is of great interest because the JPEG-2000 Standard admits only two-channel filter banks. A filter with higher number of channels can be simulated by cascading two-channel filter banks. So the algorithm has its own significantly large area of applications within the JPEG-2000 community.

4.8   Existence of Principal Component Filter Banks

4.8.1   Optimum compaction and principal component filter banks

Akkarakaran and Vaidyanathan [9] study the existence of principal component filter banks from the perspective of majorization theory and convex analysis and optimization. If it exists, a principal component filter bank renders a maximum for any convex objective function of the output sub-band variance.

Denote by $\mathcal{S}$ the output sub-band variance vectors from class $\mathcal{C} \subset \mathcal{U}_{\infty, M}$ of orthonormal filter banks and call $\mathcal{S}$ the search space. The analysis in [9] focuses on the search space to draw results on the existence of principal component filter banks for the class $\mathcal{C}$ of filter banks.

The study is an excellent theoretical tool, if a principal component filter bank exists for the class of filters considered. However, the very existence of principal component filter banks for classes of finite impulse response filter banks other than transform coders is still an unanswered question.

When principal component filter banks exist for a given class of orthonormal filter banks $\mathcal{C}$ the search space $\mathcal{S}$ has a special shape: it is a polytope, i.e. a convex set generated by a finite set of vectors in $\mathcal{S}$. Moreover, the corners of the polytope are the sub-band variance vectors that majorize all the other vectors in $\mathcal{S}$. In fact, the components of the corners are formed by permutations of one vector, corresponding to a principal component filter bank – see Remark 4.3.3.

This leads to the "sequence of compaction filters algorithm". The vector that majorizes the other vectors in the search space $\mathcal{S}$ is found as

1. Permute the vectors of $\mathcal{S}$ such that their components are ordered decreasingly.

2. Select the set S_0 of vectors such that the first component is the largest for all vectors in $\mathcal{S}$. The existence of such maximal element can be shown in a similar manner Theorem 4.6.1 is proved.

3. From S_0 select the set S_1 of vectors whose second component is the largest.

4. Continue selecting the sets S_k, $k = 2, \ldots, M - 1$ such that the $k = 1$st component is the largest.

5. The set S_{M-1} has only one element by Remark 4.3.3 and that is the output sub-and variance vector corresponding to the principal component filter bank in $\mathcal{C}$, if it exists.

The algorithm produces the maximal element with respect to the lexicographic order of $\mathbb{R}^M$. The lexicographic order is a weaker criterion for the search of a principal component filter bank. In [9] it is shown that the algorithm may be sub-optimum in the absence of principal component filter banks. In the study in Section 4.7.3 the sequence of compaction filters identifies the vector corresponding to an optimum compaction gain filter bank for which the coding gain is 3.29 dB; the maximum coding gain is 3.43 dB (see Table 4.3).

4.8.2 Discussion on the existence of principal component filter banks

The study of existence of principal component filter banks over classes of orthonormal filter banks has three main ingredients: the input signal, the number of channels and the type of filters considered.

The input. In the trivial case, for a white noise input, any orthonormal filter bank is a principal component filter bank and there is no coding gain advantage, i.e., the coding gain is 1 – see [118].

We continue with the study of nontrivial inputs. In order to have an optimal filter bank – in the coding gain sense – the blocked input power spec-

152

tral density matrix has to be diagonalized by the analysis polyphase matrix. This diagonalization is always achieved by infinite impulse response filters (the "peel-off" procedure in Theorem 3.3.3. However, in general it is not true that a finite McMillan degree polyphase matrix diagonalizes the input power spectral density matrix at each frequency. Our study with autoregressive AR(1) signals shows that filters in $\mathcal{U}_{1,3}$ do not achieve this diagonalization. When the power spectral density matrix is a polynomial, the Smith Form representation from Theorem 3.1.2 suggests that there might exist finite impulse response principal component filter banks with more than two channels. This is a possible research direction. The problem of identifying signals whose blocked power spectral density matrix is diagonalizable by finite impulse response filter banks is still open.

For orthogonal transforms, there is always an optimal transform that diagonalizes the auto-covariance input matrix namely the Karhunen-Loève transform. However, only a Gaussian input guarantees optimality of the Karhunen-Loève transform in a coding scheme.

The number of channels. Assume the input is non-trivial with integrable power spectral density. For only one channel, a filter that is optimum in the coding gain sense is also an optimum compaction, hence it is a principal component filter. The existence of such a filter is guaranteed by Theorem 4.6.1 for finite impulse response filters and by the Theorem 4.4.1 (the "peel-off" procedure) for filters with unrestricted orders.

153

1. Two channels.

We now discuss the class of filter banks with two channels. The sub-band variance vector $\boldsymbol{\sigma}_* = [\sigma_{*0}^2, \sigma_{*1}^2]^T$ that is output from the coding gain optimal filter bank guaranteed by Theorem 4.6.1 is also a principal component filter bank because $\boldsymbol{\sigma}_*$ is a maximal element in $\mathcal{S} \subsetneq \mathbb{R}_+^2$ with respect to the majorization relation. See Remark 4.3.3. In Section 4.7.2 it is shown that an optimum compaction gain filter bank is also a coding gain optimal filter bank – Equation 4.14. The coding gain of the Karhunen-Loève Transform is the lowest among the gains from principal component filter banks – and our study supports this assertion. The infinite impulse response principal component filter bank renders the maximum coding gain because it is only the "peel-off" procedure that realizes a diagonalization of the power spectral density matrix at all frequencies, equivalent to realization of the minimum value of the product of variances.

There exists a principal component filter bank in any class $\mathcal{C} \subset \mathcal{U}_{\infty,M}$ of filter that is optimal over the class $\mathcal{C}$ – but which one is *the* optimal filter bank?

From Theorem 3.3.1 it follows that the higher the filter orders the more statistics of the data are accounted for therefore a better decorrelation. This proportionality of the coding gain with respect to the filter order is also observed by Unser [115].

Note that the equivalence between majorization of sub-band spectra and that of the sub-band variances is valid only for that filter bank that is

optimal over the whole $\mathcal{U}_{\infty,M}$ and whose coding gain cannot be further increased. For two channel finite impulse response filter banks, although there is majorization of sub-band variances because the majorization relation is a total order in $\mathbb{R}_+^2$, there is no optimality in the sense of Theorem 3.3.3. By Theorem 4.6.1, in any class $\mathcal{C} \subset \mathcal{U}_{\infty,M}$ there exists a finite response filter bank that is optimal over $\mathcal{C}$ but there is no spectral majorization because the coding gain can be increased – as shown in the proof of Theorem 3.3.2. That optimal filter bank for $\mathcal{U}_{\mu,M}$ whose existence is guaranteed by Theorem 4.6.1 is also a principal component filter bank and an optimum compaction filter bank.

2. More than two channels.

Consider now finite impulse response filter banks with more than two channels. In order to attain coding gain optimality in the sense of Theorem 3.3.3, the input power spectral density matrix has to be diagonalized by a such a filter of low degree and we expect the input to be a special case. The requirements of Theorem 3.3.3 on necessary and sufficient conditions for optimality and the majorization of sub-band variances are obviously met for infinite impulse response filters constructed with the "peel-off" procedure from Proposition 4.4.1; also those filters are principal component filter banks as well. In classes of finite impulse response filter banks with more than two channels we agree with the conclusion in [62] that in general principal component filter banks with finite Macmillan degree do not exist. Our experiment for the three channel case for the model autoregressive signal supports this claim.

The filters. There are classes of filter banks definitely without prin-

cipal component filter banks like the class of discrete Fourier transforms and cosine-modulated filter banks. The filters in these classes are derived from a prototype filter and the "shape" of the filter cannot be preserved under optimality constraints. See Akkarakaran and Vaidyanathan [9].

In the class of transform coders, which are finite impulse response filter banks with zero McMillan degree, principal component filter banks always exist and they are the Karhunen-Loève Transform of the input signal, for any number of channels – under the constraint that the input is a wide sense stationary Gaussian process.

In general, for more than two channels there are no principal component filter banks unless the polyphase matrix is either constant or corresponds to an ideal passband filter bank.

When the filter orders are not restricted the principal component filter banks exist by the Theorem 3.4.1. And it is that filter bank that diagonalizes the input power spectral density matrix and renders the maximum coding gain over all filter banks in $\mathcal{U}_{\infty,M}$.

4.9 Conclusion

The problem of existence of principal component filter banks for the class $\mathcal{U}_{\mu,M}$ of M-channels finite impulse response orthonormal filter banks with McMillan degree μ is still open for $M > 2$ and $\mu > 0$.

There exist principal component filter banks for the class of memoryless

filter banks ($\mu = 0$), for the class of ideal filter banks and also for two-channel filter banks. Our contribution to the existing theory is to prove that coding gain optimal finite impulse response filter banks exist. The JPEG-2000 Standard accepts only two-channel filter banks. The algorithm we provide in Sections 4.7.4 and 4.7.2 for the construction of two-channel principal component filter banks has a wide area of application. The optimal filter banks we construct can be embedded into the JPEG-2000 for decorrelation in the component direction of multicomponent images. Also, in other applications, higher number of channels can be simulated by cascading these two-channel filter banks [117]. The filters we propose are finite impulse response hence they lead to compactly supported wavelets [31, 32, 24].

Our design of two-channel filter banks has the advantage that is simple and does not resort to any model of the input data. The area of applicability makes it valuable as a decorrelation tool for large collections of images.

From Equations (4.9) and (4.10) it follows that the Karhunen-Loève transform which is the principal component filter bank in $\mathcal{U}_{0,M}$ and the principal component filter bank in $\mathcal{U}_{1,M}$ do not differ by just a delay term. Therefore, it is not possible to construct principal component filter banks by adding memory to lower degree filter banks known to be optimal.

A suitable approach would be to find non-trivial classes of input statistics for which there exist principal component filter banks. One decorrelation method not yet deeply studied is via independent component analysis [64], which overcomes first of all the constraint on the input to be Gaussian [51] for

157

the class of orthogonal transforms.

Chapter 5

Application to Remote Sensing Imagery

5.1 Introduction

In this last part of the thesis we show how the theoretical results presented in the previous chapters are employed in the processing and analysis of hyper-spectral images.

A high level overview of the JPEG-2000 Standard and a description the data used in experiments are first presented. We then show the performance of transform coders and filter banks with McMillan degree 1. These filter banks are principal component filter banks.

Pattern recognition and data exploitation techniques need not limit remote sensing systems to lossless compression. Many common feature extraction tasks are reliable on data that is compressed at very low bit rates [86].

We conclude with a summary of results and present directions for future research.

5.2 The JPEG-2000 Standard

The new still image compression standard, JPEG-2000, is written with explicit requirements for supporting coding and compression of multicomponent imagery. The standard is equipped with a variety of options for transforming and manipulating collections of image components (e.g., spectral bands). A high level overview of the JPEG-2000 Standard is presented in Figure 5.1. Component decorrelation transforms are followed by spatial wavelet transforms, rate allocation and quantization, and binary arithmetic bit-plane encoding. The compressed bit-streams are signaled in packets that can be ordered according to a variety of possible priorities to support various progressive transmission objectives. For instance, the initial portion of a truncated JPEG-2000 code-stream can always be decoded to yield an approximation of the image represented by the full code-stream. This feature, known as an *embedded code-stream,* enables multiple users to access the same compressed code-stream at a variety of different levels of fidelity, which is useful in database applications. A detailed technical presentation of the JPEG-2000 standard can be found in Taubman and Marcellin [109]. In the experiments below, an AVIRIS hyperspectral image – described in Section 5.3 – is compressed and reconstructed using JPEG-2000 at bit rates varying from 0.125 bits/pixel/band (bpppb) to 4 bpppb. Equivalently, the compression ratios vary from 4 to 128 (the data is recorded at 16 bits/pixel/band). AVIRIS data is highly correlated along the spectral axis, and we exploit this fact with 2 different component decorrelation transforms. The fidelity of reconstructed images is quantified and reported as

Source data cube

```
x  →  Component      →  Spatial       →  Scalar         →  Binary
       Decorrelation      Wavelet          Quantization      Arithmetic
       Transform          Transform        Encoding          Encoding
```

Compressed
information

```
x̂  ←  Inverse       ←  Inverse       ←  Scalar         ←  Binary
       Component         Wavelet          Quantization      Arithmetic
       Transform         Transform        Decoding          Decoding
```
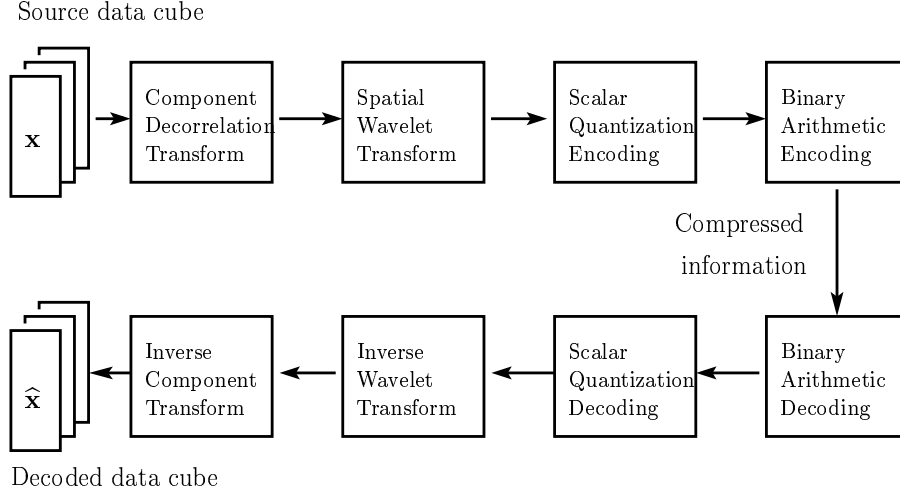
Decoded data cube

Figure 5.1: High level overview of JPEG-2000.

a function of bit rate.

For instance, one simple way of quantifying fidelity is to report the 3-dimensional signal to noise ratio of a reconstructed image cube ("3-D SNR"), where "truth" is given by the original, uncompressed image cube. Typical rate-distortion performance for such measurements is shown in Figure 5.2. The curves present 3-dimensional signal to noise ratio as a function of the bit rate for 3 component transform options: no component decorrelation, wavelet transform decorrelation, and Karhunen-Loève decorrelation (denoted "KLT" in the figure). The other steps in the JPEG-2000 process were identical in all 3 cases. Observe how (non-adaptive) 9-7 wavelet transform decorrelation yields a gain of around 10-12 dB on this particular image, while image-dependent decorrelation produces around 15-20 dB of gain.
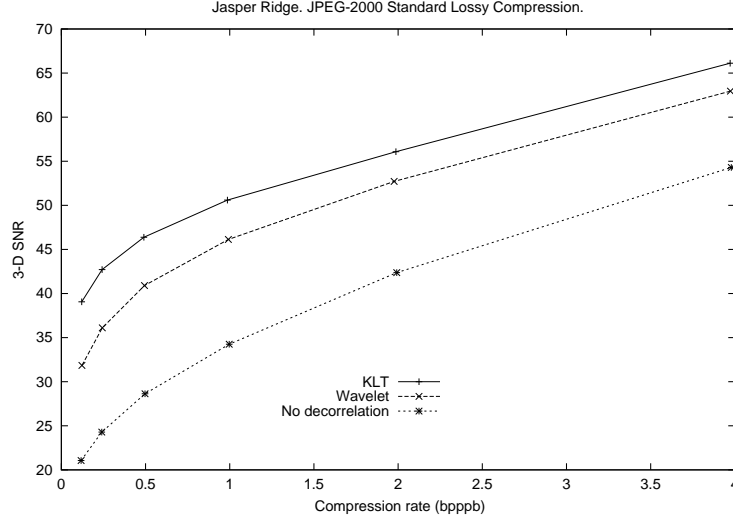
161

Figure 5.2: 3-D SNR for the "Jasper Ridge" image.

5.3 Data Used in Experiments

The experiments are performed on Airborne Visible/InfraRed Imaging Spectrometer (AVIRIS) hyper-spectral images. AVIRIS is a Jet Propulsion Laboratory instrument that delivers calibrated images of the upwelling spectral radiance in 224 contiguous spectral bands with wavelengths from 0.37340 μm to 2.50326 μm with a spatial resolution varying from a few meters to 20 meters. Each spectral component has 512 x 614 pixels with a sample precision of 16 bpppb [123, 122].

The "Jasper Ridge Scene" (Figure 5.3) is an image of a biological preserve in California and it used in studies for monitoring seasonal vegetation patterns.

The high correlation of hyper-spectral images allows us to assume wide sense stationarity of these signals. In our applications we block them by a

162

factor of 224 which is the number of spectral components.

5.4   Design of Two-channel Optimal Filter Banks

We seek the design of a two-channel filter bank with McMillan degree 1 that is optimal in the coding gain sense – and in the principal component sense. Recall that any class $\mathcal{U}_{\mu,2}$ has principal component filter banks. We also determine the optimum orthogonal transform and compare the coding gains for filter banks with McMillan degree 0 (transform coders) and 1.

We depict a hyper-spectral cube from the Jasper Ridge scene. Denote by r the number of rows, c number of columns of each two-dimensional image and by b the number of spectral bands. We select

$$r = 35, \ c = 35, \ b = 224 \,.$$

The discrete signal for which we construct a filter banks has length

$$N = r \times c \times b = 277400 \,.$$

This signal is input to a two-channel signal-adapted principal component filter bank.

The performance of this filter banks is expressed in terms of the output variances v_0 and v_1. Recall the Equations (3.7) and (3.8) for the polyphase analysis matrix:

$$
\begin{aligned}
\boldsymbol{E}(z) &= \boldsymbol{V}(z)\,\boldsymbol{U} \\
\boldsymbol{V}(z) &= \boldsymbol{I} + (-1 + z^{-1})\boldsymbol{v}\boldsymbol{v}^T, \qquad \|\boldsymbol{v}\| = 1,
\end{aligned}
$$

Figure 5.3: The Jasper Ridge scene

where

$$\boldsymbol{v} \;=\; [\cos \alpha \;\; \sin \alpha]^T, \qquad \alpha \in [0, 2\pi],$$

$$\boldsymbol{U} \;=\; \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix}.$$

as in Section 4.7.2. The filters $H_0(\cdot)$, $H_1(\cdot)$ and the polyphase matrix are related as

$$\begin{bmatrix} H_0(z) \\ H_1(z) \end{bmatrix} = \boldsymbol{E}(z^2) \begin{bmatrix} 1 \\ z^{-1} \end{bmatrix}.$$

The polyphase matrix for the class of orthogonal transforms is

$$\boldsymbol{E}(z) = \boldsymbol{U} = \text{constant}$$

with $\boldsymbol{U}$ as previously defined.

The coding gain for both cases, orthogonal transform and filter banks with memory is

$$G_{coding} = \frac{\sigma_{v_0}^2 + \sigma_{v_1}^2}{2\sqrt{\sigma_{v_0}^2\,\sigma_{v_1}^2}}$$

164

where $\sigma_{v_0}^2$, $\sigma_{v_1}^2$ are the output sub-band variances calculated as (see Section 1.2.2)

$$\sigma_{v_k}^2 = \frac{1}{2\pi} \int_0^{2\pi} S_{xx}(e^{i\omega})|H_k(e^{i\omega})|^2 \, d\omega \, .$$

We estimate the power spectral density of the input signal using the Matlab periodogram. The expressions of the filters are also generated using symbolic computation and the output variances are calculated numerically using the formula shown above. We estimate the power spectral density for the whole signal of length $N = 277400$ and also by considering the signal a vector random process of dimension 224, the number of spectral bands. The values of the output sub-band variances are the same for both estimation method and this is explained by the wide-sense stationarity of the hyper-spectral images.

We find that the optimal orthogonal transform with respect to the coding gain is attained for $\boldsymbol{U}(\theta) \cong \pi/4$. This is the Karhunen-Loève transform of a wide-sense stationary process – see Section 4.7.2. The maximum coding gain for the class $\mathcal{U}_{0,2}$ of orthogonal transforms is 8.9684dB. For filter banks with McMillan degree 1 the highest coding gain value of 11.3025 dB is rendered by a filter described by the parameters $\alpha = 1.83$, $\theta = 2.363$. From the optimal orthogonal transform to the optimal McMillan degree one filter bank there is a net substantial increase in the coding gain of 2.334 dB.

5.5   Future Work

Principal component filter banks unify the theory on the optimality of filter banks under explicitly stated criteria. Finite impulse response principal

165

component filter banks do not exist for general classes of input; however this remains an outstanding issue.

We address the issue of optimal coding gain filter banks and show that in any finite impulse response class of orthonormal filter banks there exists a coding gain optimal filter bank. On the practical side we provide an algorithm for the design of coding gain optimal finite impulse response filter banks that are signal-adapted for a large class of multicomponent images.

Coding gain optimality is a widely used metric for the performance of a filter bank; for classes of filter banks and input signals without known principal component filter banks, coding gain optimal filter banks are a good, practical alternative. When the number of channels is two, these optimal signal-adapted filter banks are also principal component filter banks; they can be embedded into the JPEG-2000 Standard as a decorrelation tool, leading to an efficient compression.

Our future work will address theoretical issues and applications. A few possible directions are:

- Compare the performance of these filter banks with the performance of the Karhunen-Loève transform for all 224 channels (spectral bands). The cost of implementing a two-channel filter bank is significantly lower than that for the 224-channel Karhunen-Loève transform which is.

- Address the implementation issue of these two-channel filter banks for their embedding into the JPEG-2000 Standard. That involves the design

of the lifting steps for the filter implementation [24, 13, 108] and full automatizing of the process. In other words, it is desired to have as input only the multicomponent image and design a black box to determine the input statistics, two-channel principal component filter bank and lifting steps and to filter the image.

- Research on the existence of particular classes classes of input that have finite impulse response principal component filter banks. In particular address this issue when the input power spectral density matrix is – or can be approximated by – a low order polynomial so that a finite impulse response polyphase matrix diagonalizes this polynomial matrix [28, 76, 45].

- Perform a comprehensive statistical approach to the existence of classes of input signals and filter banks for which principal component filter banks exist.

- Perform pattern recognition and data exploitation studies on decorrelated (transformed) and compressed/reconstructed images using adaptive principal component filter banks.

Our contribution to the theory of existence of coding gain optimal finite impulse response filter banks is a response to the quest for optimal filter banks for classes of filter banks that do not have principal component ones. Theoretical studies with statistical tools will open new research directions in the

167

existence and design of principal component filter banks and we will pursue that path.

Bibliography

[1] Visible human project web site. http://www.nlm.nih.gov/research/visible.

[2] *JPEG 2000 Image Coding System, Part 1*, ISO/IEC Int'l. Standard 15444-1, ITU-T Rec. T.800. Int'l. Org. for Standardization, December 2000.

[3] *JPEG 2000 Image Coding System, Part 2 (Extensions)*, ISO/IEC Int'l. Standard 15444-2, ITU-T Rec. T.800. Int'l. Org. for Standardization, December 2001.

[4] K. C. Aas, K. A. Duell, and C. T. Mullis. Synthesis of extremal wavelet-generating filters using Gaussian quadrature. *IEEE Transactions on Signal Processing*, 43(5):1045–1057, 1995.

[5] H. Abut. *Vector Quantization*. IEEE Press, New York, 1990.

[6] S. Akkarakaran and P. P. Vaidyanathan. The best basis problem, compaction problem and principal component filter bank design problems. In *Proc. IEEE Int. Symp. Circuits and Systems*, volume 3, pages 508–511, Orlando, FL, 1999.

[7] S. Akkarakaran and P. P. Vaidyanathan. On optimization of filter banks with denoising applications. In *Proc. IEEE International Symposium on Circuits and Systems*, volume 3, pages 512 – 515, Orlando, FL, 1999.

[8] S. Akkarakaran and P. P. Vaidyanathan. The role of principal component filter banks in noise reduction. In *Proceedings of SPIE – The International Society for Optical Engineering*, volume 3813, pages 346–357, 1999.

[9] S. Akkarakaran and P. P. Vaidyanathan. Filter bank optimization with convex objectives and the optimality of principal component forms. *IEEE Trans. Signal Processing*, 49(1):100–114, 2001.

[10] S. Akkarakaran and P. P. Vaidyanathan. A review of the theory and applications of optimal subband and transform coders. *Applied and Computational Harmonic Analysis*, (10):254–289, 2001.

[11] A. Al-Smadi and D. M. Wilkes. Robust and accurate ARX and ARMA model order estimation of non-Gaussian processes. *IEEE Transactions on Signal Processing*, 30(3):759–763, 2002.

[12] V. R. Algazi and D. J. Sakrison. On the optimality of the Karhunen-Loève expansion. *IEEE Trans. on Information Theory*, 15:319–321, 1969.

[13] A. Z. Averbuch and V. A. Zheludev. Lifting scheme for biorthogonal multiwavelets originated from Hermite splines. *IEEE Transactions on Signal Processing*, 50(3):487–500, 2002.

[14] M. Bellanger. *Digital processing of signals: theory and practice.* John Wiley and Sons, New York, third edition, 2000.

[15] R. Bellman. *Introduction to Matrix Analysis.* McGraw-Hill, New York, second edition, 1970.

[16] K. Berberidis and S. Theodoridis. Efficient symmetric algorithms for the modified covariance method for autoregressive spectral analysis. *IEEE Trans. on Signal Processing*, 41(1):43–54, 1993.

[17] D. R. Brillinger. *Time series and data analysis.* Holt, Rinehart and Winston, New York, 1975.

[18] C. M. Brislawn. Kernels of trace class operators. In *Proceedings of the American Mathematical Society*, volume 104, pages 1181–1190, 1988.

[19] C. M. Brislawn. Rational transfer matrices with FIR inverses. *IEEE Signal Process. Soc., Proc. Int'l. Symp. Time-Freq. Time-Scale Analysis, (Paris, France)*, pages 53–56, 1996.

[20] C. M. Brislawn, J. N. Bradley, R. J. Onyshczak, and T. Hopper. The FBI compression standard for digitized fingerprint images. *Proc. SPIE Appl. Digital Image Process. XIX (Denver, CO)*, 2847:344–355, August 1996.

[21] C. M. Brislawn, W. B. Clodius, N. R. Harvey, M. D. Quirk, and J. Theiler. Multispectral and hyperspectral image processing. In R. G. Driggers, editor, *Part 3. Transforms, Classification and Coding*, Encyclopedia of Optical Engineering, To appear. Marcel Dekker, Inc., New York, NY, 2003.

171

[22] C. M. Brislawn and M. D. Quirk. *Image Compression with the JPEG-2000 Standard.* Encyclopedia of Optical Engineering, To appear. Marcel Dekker, Inc., New York, NY, 2003.

[23] B. V. Brower, A. Lan, J. Kasner, and S. Shen. Multiple component compression within JPEG-2000 as compared with other techniques. *Applications of Digital Image Processing XXIII, Proceedings of SPIE*, 4115:544–551, 2000.

[24] R. Calderbank, I. Daubechies, W. Sweldens, and B.-L. Yeo. Wavelet transforms that map integers to integers. *Applied and Computational Harmonic Analysis*, 5(3):332–369, 1998.

[25] W. Cheney. *Analysis for Applied Mathematics.* Springer-Verlag, New York, 2001.

[26] W. Cheney and W. Light. *A Course in Approximation Theory.* Brooks Cole, Pacific Grove, CA, 1999.

[27] W. B. Clodius, P. G. Weber, C. C. Borel, and B. W. Smith. Multispectral band selection for satellite based systems. In Gerald C. Holst, editor, *Infrared Imaging Systems: Design, Analysis, Modeling, and Testing IX*, volume 3377 of *Proc. SPIE*, pages 11–21, 1998.

[28] J. E. Cohen, J. H. B. Kemperman, and G. Zbaganu. *Comparisons of Stochastic Matrices with Applications in Information Theory, Statistics Economics and Population Sciences.* Birkhauser, Boston, 1998.

[29] R. R. Coifman, Y. Meyer, and M. V. Wickerhauser. Wavelets and their applications. In Ruskai et al., editor, *Wavelet Analysis and Signal Processing*, pages 153–178. Jones and Barlett, Boston, 1992.

[30] R. E. Crochiere and L. R. Rabiner. Interpolation and decimation of digital signals: a tutorial review. *Proc. IEEE*, 69:300–331, 1981.

[31] I. Daubechies. Orthogonal bases of compactly supported wavelets. *Comm. Pure Appl. Math*, 41:909–996, 1988.

[32] I. Daubechies. *Ten Lectures on Wavelets*. SIAM, 1992.

[33] W. B. Davenport and W. L. Root. *An Introduction to the Thoery of Random Signals and Noise*. IEEE Press, The Institute of Electrical and Electronics Engineers, Inc., New York, 1987. Reprint of a book published by McGraw Hill Book Company in 1958 under the same title.

[34] C. E. Davila. Blind adaptive estimation of KLT basis vectors. *IEEE Transactions on Signal Processing*, 49(7):1364–1369, 2001.

[35] O. A. de Carvalho, Jr. and P. R. Meneses. Spectral correlation mapper (SCM): An improvement on the spectral angle mapper (SAM). In *Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) 2000 Workshop Proceedings*, Pasadena, 2000. JPL NASA.

[36] P. Delsarte, B. Macq, and D. T. Slock. Signal adapted multiresolution transform for image coding. *IEEE Trans. Information Theory*, 38:897–904, 1992.

[37] J. L. Doob. *Stochastic Processes.* John Wiley and Sons, New York, 1953.

[38] W. Feller. *An Introduction to Probability Theory and Its Applications.* John Wiley and Sons, 3rd edition, 1968.

[39] H. Feng and M. Effros. On the rate-distortion performance and computational efficiency of the Karhunen-Loève transform for lossy data compression. *IEEE Transactions on Image Processing*, 11(2):113–122, 2002.

[40] A. Gersho and R. M. Gray. *Vector quantization and signal compression.* Kluwer Academic Publisher, Boston, 1991.

[41] J. Gilbert. *Notes on Wavelets.* University of Texas at Austin, 1997.

[42] I. Gohberg, P. Lancaster, and L. Rodman. *Matrix Polynomials.* Academic Press, New York, 1982.

[43] G. H. Golub and C. F. Van Loan. *Matrix Computations.* John Hopkins University Press, Baltimore and London, 3rd edition, 1996.

[44] R. M. Gray. *Non-US data compression and coding research.* Science Applications International Corporation, McLean, VA, 1993.

[45] L. Gurvits. Factorization of positive matrix polynomial and the separability of composite quantum systems. Technical Report LA-UR-01-2030, Los Alamos National Laboratory, 2001.

[46] G. H. Hardy, J. E. Littlewood, and G. Polya. *Inequalities.* Cambridge University Press, London and New York, 2nd edition, 1952.

[47] R. A. Horn and C. R. Johnson. *Matrix Analysis.* Cambridge University Press, Cambridge, United Kingdom, 1985.

[48] R. A. Horn and C. R. Johnson. *Topics in Matrix Analysis.* Cambridge University Press, Cambridge, United Kingdom, 1991.

[49] H. Hotelling. Analysis of a complex of statistical variables into principal components. *J. Educ. Psychology*, 24(1933):417–441 and 498–520.

[50] H. Hotelling. Relations between two sets of variates. *Biometrika*, 28:321–377, 1936.

[51] Y. Huang and P. M. Schultheiss. Block quantization of correlated Gaussian random variables. *IEEE Trans. Commun. Systems*, C-10:298–296, 1963.

[52] A. Hyvarinen, J. Karhunen, and E. Oja. *Independent Component Analysis.* John Wiley and Sons, Inc., New York, 2001.

[53] N. Jacobson. *Basic Algebra I.* W. H. Freeman & Co., New York, 1974.

[54] A. K. Jain. *Fundamentals of Digital Image Processing.* Prentice-Hall, Englewood Cliffs, NJ, 1989.

[55] N. J. Jayant, J. B. Johnstone, and B. Safranek. Signal compression based on models of human perception. *Proc. IEEE*, 81(10):1385–1422, 1993.

[56] N. S. Jayant and P. Noll. *Digital Coding of Waveform*. Prentice Hall, Englewood Cliffs, NJ, 1984.

[57] R. E. Kalman. Irreducible realizations and the degree of a rational matrix. *SIAM J. Appl. Math.*, 13:520–544, 1965.

[58] H. Karhunen. *Uber Lineare Methoden in der Wahrscheinlich-Keitsrechnung*. Ann. Acad. Science Fenn, Ser.A.I,37, Helsinki, 1947. (see translation by I. Selin in the Rand Corp., Doc T131, August 11, 1960).

[59] J. H. Kasner, A. Bilgin, M. W. Marcellin, A. Lan, B. V. Brower, S. S. Shen, and T. S. Wilkinson. JPEG-2000 compression using 3D wavelets and KLT with application to HYDICE data. *Imaging Spectrometry VI, Proceedings of SPIE*, 4132:157–166, 2000.

[60] D. Kincaid and W. Cheney. *Numerical Analysis – Mathematics of Scientific Computing*. Brooks Cole, Pacific Grove, CA, second edition, 1990.

[61] A. Kirac. *Optimal Orthonormal Subband Coding and Lattice Quantization with Vector Dithering*. PhD thesis, California Institute of Technology, Pasadena, CA, 1998.

[62] A. Kirac and P. P. Vaidyanathan. On existence of FIR principal compo-nent filter banks. In *Proc. IEEE ICASSP*, volume 3, pages 1329–1332, Seattle, WA, 1998.

[63] A. Kirac and P. P. Vaidyanathan. Theory and design of optimum FIR compaction filters. *IEEE Transactions on Signal Processing*, 46(4):903–919, 1998.

[64] T. W. Lee. *Independent Component Analysis: Theory and Applications.* Kluwer Academic Publishers, Boston, 1998.

[65] A. Levy and M. Lindenbaum. Sequential Karhunen–Loeve basis ex-traction and its application to images. *IEEE Transactions on Image Processing*, 9(8):1371–1374, 2000.

[66] S. P. Lloyd. *Least Squares Quantization in PCM's.* Bell Telephone Laboratories Paper, Murray Hill, 1957.

[67] S. P. Lloyd. A sampling theorem for stationary (wide sense) stochastic processes. *Trans. Amer. Math. Soc.*, (921):1–12, 1959.

[68] M. Loeve. Fonctions aleatoires de seconde ordre. In P. Levi, editor, *Pro-cessus Stochastiques et Mouvement Brownien*. Hermann, Paris, France, 1948.

[69] G. Lohmann. *Volumetric Image Analysis.* John Wiley and Sons, New York, 1998.

177

[70] N. Lu. *Fractal Imaging.* Academic Press, San Diego, 1997.

[71] S. Mallat. A theory of multiresolution signal decomposition: The wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11:674–693, 1989.

[72] S. Mallat. *A Wavelet Tour of Signal Processing.* Academic Press, San Diego, 2nd edition, 1999.

[73] A. W. Marshall and I. Olkin. *Inequalities: Theory of Majorization and Its Applications.* Academic Press, San Diego, 1979.

[74] J. Max. Quantizing for minimum distortion. *IRE Trans. Inform. Theory*, IT-6:7–12, 1960.

[75] A. Mertins. *Signal Analysis: Wavelets, Filter Banks, Time-Frequency Transforms and Applications.* John Wiley and Sons, New York, reprinted edition, 2000.

[76] K. S. Miller. *Some Eclectic Matrix Theory.* Krieger, Malabar, FL, 1987.

[77] G. Mirchandani, R. Foote, D. N. Rockmore, D. Healy, and T. Olson. A wreath product group approach to signal and image processing – part II: Convolution, correlation and applications. *IEEE Transactions on Signal Processing*, 48(3):749 – 767, 2000.

[78] B. Moghaddam and A. Pentland. Probabilistic visual learning for object representation. *IEEE Trans. on Pattern Analysis and Machine Intelligence;*, 19(7):696–710, 1997.

[79] P. Moulin, M. Anitescu, K. O. Kortanek, and F. Potra. The role of linear semi–infinite programming in signal–adapted QMF bank design. *IEEE Transactions on Signal Processing,*, 45(9):2160–2174, 1997.

[80] P. Moulin and M. K. Mihcak. Theory and design of signal-adapted FIR paraunitary filter banks. *IEEE Transactions on Signal Processing*, 46(4):920 – 929, 1998.

[81] H. Murase and M. Lindenbaum. Partial eigenvalue decomposition of large images using spatial temporal adaptive method. *IEEE Transactions on Image Processing*, 4(5):620 – 629, 1995.

[82] H. Murase and S. K. Nayar. Visual learning and recognition of 3-D objects from appearance. *International Journal of Computer Vision*, 14(1):5–24, 1995.

[83] F. D. Murnaghan. *The Unitary and Rotation Groups*. Spartan Books, Washington, D.C., 1962.

[84] A. V. Oppenheim and R. W. Schafer. *Digital signal processing*. Prentice Hall, Englewood Cliffs, NJ, 1975.

[85] A. V. Oppenheim and R. W. Schafer. *Discrete-time signal processing*. Prentice Hall, Englewood Cliffs, NJ, 1989.

[86] M. D. Pal, C. M. Brislawn, and S. P. Brumby. Feature extraction from hyperspectral images compressed using the JPEG-2000 Standard.

pages 168–172, Sante Fe, NM, 2002. Fifth IEEE Southwest Symposium on Image Analysis and Interpretation.

[87] B. P. Palka. *An Introduction to Complex Function Theory.* Springer-Verlag, New York, 1991.

[88] A. Papoulis. *Probability, Random Variables and Stochastic Processes.* McGraw-Hill, Boston, MA, 3rd edition, 1991.

[89] W. B. Pennebaker and J. L. Mitchell. *JPEG: Still Image Data Compression Standard.* Chapman and Hall, New York, 1993.

[90] M. D. Quirk. New results on optimal filter banks. Technical Report LA-UR-02-6361, Los Alamos National Laboratory, 2002.

[91] W. D. Ray and R. M. Driver. Further decomposition of the Karhunen-Loève series representation of a stationary random process. *IEEE Trans. Inform. Theory*, (IT-16):663–668, 1970.

[92] J. A. Richards and X. Jia. *Remote Sensing Digital Image Analysis.* Springer-Verlag, Berlin Heidelberg, third edition, 1999.

[93] F. Riesz and B. Sz.-Nagy. *Functional analysis.* Dover Publications, Inc., New York, reprint edition, 1990.

[94] S. Roman. *Introduction to Signal Coding and Information Theory.* Springer-Verlag, New York, 1997.

[95] W. Rudin. *Principles of Mathematical Analysis*. McGraw-Hill, New York, 3rd edition, 1976.

[96] W. Rudin. *Real and Complex Analysis*. McGraw-Hill, New York, 3rd edition, 1987.

[97] V. P. Sathe and P. P. Vaidyanathan. Effects of multirate systems on the statistical properties of random signals. *IEEE Trans. on Signal Processing*, 41(1):131–146, 1993.

[98] P. Schelkens, J. Barbarien, and J. Cornelis. Compression of volumetric medical data based on cube-splitting. In *Proceedings of SPIE - The International Society for Optical Engineering*, volume 4115, pages 91–101, 2000.

[99] R. A. Schowengerdt. *Remote Sensing Models and Methods for Image Processing*. Academic Press, San Diego, second edition, 1997.

[100] A. Segall. Bit allocation and encoding for vector sources. *IEEE Transactions on Information Theory*, IT–22:162–169, 1976.

[101] K. S. Shanmugan and A. M. Breipohl. *Random Signals: Detection, Estimation and Data Analysis*. John Wiley and Sons, New York, 1988.

[102] S. S. Shen and K. H. Kasner. Effects of 3D wavelets and KLT based JPEG-2000 - hyperspectral compression on exploitation. *Imaging Spectrometry VI, Proceedings of SPIE*, 4132:167–176, 2000.

181

[103] L. Sirovich and M. Kirby. Low-dimensional procedure for the characterization of human faces. *Journal of the Optical Society of America A (Optics and Image Science);*, 4(3):519–524, 1987.

[104] V. Smirnov. *Cours de Mathematiques Superieures.* Editions MIR, Moscou, 1975.

[105] A. K. Soman and P. P. Vaidyanathan. On orthonormal wavelets and paraunitary filter banks. *IEEE Trans. on Signal Processing*, 41(3):1170–1183, 1993.

[106] G. Strang and T. Nguyen. *Wavelets and Filter Banks.* Wellesley-Cambridge Press, Wellesley, MA, 1996.

[107] B. W. Suter. *Multirate and wavelet signal processing.* Academic Press, San Diego, 1998.

[108] P. Sweldens and R. Piessens. Quadrature formulae and asympotic error expansions for wavelet approximations of smooth functions. *SIAM J. Numerical Analysis*, 31:377–401, 1994.

[109] D. S. Taubman and M. W. Marcellin. *JPEG–2000: Image Compression Fundamentals, Standards, and Practice.* Kluwer Academic Publishers, Boston, 2001.

[110] T. D. Tran, R. L. deQueiroz, and T. Q. Nguyen. Linear-phase PR filter banks: Lattice structure, design and applications in image coding. *IEEE Transactions on Signal Processing*, 48(1):133 – 147, 2000.

[111] T. D. Tran, M. Ikehara, and T. Q. Nguyen. Linear-phase paraunitary filter banks with filters with different lengths and its applications in image compression. *IEEE Transactions on Signal Processing*, 47(10):2730 – 2744, 1999.

[112] H. L. Van Trees. *Detection, Estimation and Modulation Theory, Part I.* John Wiley and Sons, New York, 1968.

[113] M. K. Tsatsanis and G. B. Giannakis. Principal component filter banks for optimal multiresolution analysis. *IEEE Trans. on Signal Processing*, 43(8):1766–1777, 1995.

[114] M. Unser. An extension of the Karhunen-Loève transform for wavelets and perfect reconstruction filter banks. *Proc. Intl. Soc. for Optical Eng. (SPIE) Mathematical Imaging*, 2034:45–56, 1993.

[115] M. Unser. On the optimality of ideal filters for pyramid and wavelet signal approximation. *IEEE Transactions on Signal Processing*, 41(12):3591–3596, 1993.

[116] M. Unser. Splines: A perfect fit for signal and image processing. *IEEE Transactions on Signal Processing*, 47(11):22–38, 1999.

[117] P. P. Vaidyanathan. *Multirate Systems and Filter Banks.* Prentice-Hall, Englewood Cliffs, NJ, 1993.

[118] P. P. Vaidyanathan. Theory of optimal orthonormal subband coders. *IEEE Transactions on Signal Processing*, 46(6):1528 – 1545, 1998.

[119] P. P. Vaidyanathan and S. Akkarakaran. A review of the theory and applications of optimal subband and transform coders. *Applied and Computational Harmonic Analysis*, 10:254–289, 2001.

[120] P. P. Vaidyanathan and A. Kirac. Results on optimal biorthogonal filter banks. *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, 45(8):932–947, 1998.

[121] P. P. Vaidyanathan, T. Q. Nguyen, Z. Doganata, and T. Saramaki. Improved techniques for design of PR FIR QMF banks with lossless polyphase matrices. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 37(7):1042 – 1056, 1989.

[122] G. Vane. Airborne Visible/Infrared Imaging Spectrometer (AVIRIS). Technical Report JPL Pub. 87-38, Jet Propulsion Laboratory, Pasadena, CA, 1987.

[123] G. Vane, A. F. H. Goetz, and J. B. Wellman. Airborne imaging spectrometer: A new tool for remote sensing. *Proc. 1984 IEEE Transactions on Geoscience and Remote Sensing*, GE-22(6):546–549, 1984.

[124] M. Vetterli and C. Herley. Wavelets and filter banks: Theory and design. *IEEE Transactions on Signal Processing*, 40(9):2207–2232, 1992.

[125] M. Vetterli and J. Kovacevic. *Wavelets and Subband Coding.* Prentice Hall, Upper Saddle River, NJ, 1995.

[126] D. O. Walsh, A. F. Gmitro, and M. W. Marcellin. Adaptive reconstruction of phased array MR imagery. *Magnetic Resonance in Medecine*, 43(5):682–690, 2000.

[127] M. V. Wickerhauser. *Adapted Wavelet Analysis from Theory to Software*. A. K. Peters, Wellesley, MA, 1994.

[128] B. Wohlberg. *Fractal Image Compression and the Self-Affinity Assumption: A Stochastic Signal Modelling Perspective*. PhD thesis, University of Cape Town, Cape Town, South Africa, 1996.

[129] J. Xin and Y-Y. Qi. A perception and PDE based nonlinear transformation for processing spoken words. *Physica*, D 149:143–160, 2001.

[130] B. Xuan and R. H. Bamberger. Complete FIR principal component filter bank's. *Proc. ISCAS, Atlanta, GA*, pages 417–420, 1996.

[131] B. Xuan and R. H. Bamberger. FIR principal component filter banks. *IEEE Transactions on Signal Processing*, 46(4):930–940, 1998.

185

Vita

Mihaela Dobre Pal was born in Bucharest, Romania on November 13, 1959, the daughter of Marin Dobre and Maria Dobre. From 1984 to 1991 she worked as a programmer analyst for the Computer Manufacturing and Maintenance Company in Bucharest, Romania and from 1991 to 1994 as a software engineer for Western Geophysical Company in Houston, Texas. She received her Bachelor of Science degree in Mathematics from the University of Bucharest, Romania, in 1984 and her Master of Science degree in Computational and Applied Mathematics from the University of Texas at Austin in 1998. She joined the graduate program of the Mathematics Department at the University of Texas at Austin in 1998. She is working as a graduate research assistant at the Los Alamos National Laboratory since 1998.

Permanent address: 1860 Camino Redondo, Los Alamos, NM 87544

This dissertation was typeset with LaTeX† by the author.

---

†LaTeX is a document preparation system developed by Leslie Lamport as a special version of Donald Knuth's TeX Program.